NOVEL TELOMERASE

The present application is a continuation of U.S. Patent Application No. 09/843,676, filed April 26, 2001, which is a continuation of U.S. Patent Application No. 08/854,050, filed May 9, 1997, now U.S. Patent No. 6,261,836; which is a continuation-in-part of U.S. Patent Application No. 08/851,843, filed May 6, 1997, now U.S. Patent 6,093,809; which is a continuation-in-part of U.S. Patent Application No. 08/846,017, filed April 25, 1997, now abandoned; which is a continuation-in-part of U.S. Patent Application No. 08/844,419, filed April 18, 1997, now abandoned; which is a continuation-in-part of U.S. Patent Application No. 08/724,643, filed October 1, 1996, now abandoned. Each of the aforementioned applications is explicitly incorporated herein by reference in its entirety and for all purposes.

FIELD OF THE INVENTION

The present invention is related to novel telomerase genes and proteins. In particular, the present invention is directed to a telomerase isolated from *Euplotes aediculatus*, the two polypeptide subunits of this telomerase, as well as sequences of the *Schizosaccharomyces*, *Tetrahymena*, and human homologs of the *E. aediculatus* telomerase.

BACKGROUND OF THE INVENTION

Telomeres, the protein-DNA structures physically located on the ends of the eukaryotic organisms, are required for chromosome stability and are involved in chromosomal organization within the nucleus (*See e.g.,* Zakian, Science 270:1601 [1995]; Blackburn and Gall, J. Mol. Biol., 120:33 [1978]; Oka *et al.,* Gene 10:301 [1980]; and Klobutcher *et al.,* Proc. Natl. Acad. Sci., 78:3015 [1981]). Telomeres are believed to be essential in such organisms as yeasts and probably most other eukaryotes, as they allow cells to distinguish intact from broken chromosomes, protect chromosomes from degradation, and act as substrates for novel replication mechanisms. Telomeres are generally replicated in a complex, cell cycle and developmentally regulated, manner by "telomerase," a telomere-specific DNA polymerase. However, telomerase-independent means for telomere maintenance have been described. In recent years, much attention has been focused on telomeres, as telomere loss has been associated with chromosomal changes such as those that occur in cancer and aging.

Telomeric DNA

In most organisms, telomeric DNA has been reported to consist of a tandem array of very simple sequences, which in many cases are short and precise. Typically, telomeres consist of simple repetitive sequences rich in G residues in the strand that runs 5' to 3' toward the chromosomal end. For example, telomeric DNA in *Tetrahymena* is comprised of sequence T_2G_4, while in *Oxytricha*, the sequence is T_4G_4, and in humans the sequence is T_2AG_3 (*See e.g.,* Zakian, Science 270:1601 [1995]; and Lingner *et al.,* Genes Develop., 8:1984 [1994]). However, heterogenous telomeric sequences have been reported in some organisms (*e.g.,* the sequence $TG_{1\text{-}3}$ in *Saccharomyces*). In addition, the repeated telomeric sequence in some organisms is much longer, such as the 25 base pair sequence of *Kluyveromyces lactis*. Moreover, the telomeric structure of some organisms is completely different. For example, the telomeres of *Drosophila* are comprised of a transposable element (*See*, Biessman *et al.,* Cell 61:663 [1990]; and F.-m Sheen and Levis, Proc. Natl. Acad. Sci., 91:12510 [1994]).

The telomeric DNA sequences of many organisms have been determined (*See e.g.,* Zakian, Science 270:1601 [1995]). However, it has been noted that as more telomeric sequences become known, it is becoming increasingly difficult to identify even a loose consensus sequence to describe them (Zakian, *supra*). Furthermore, it is known that the average amount of telomeric DNA varies between organisms. For example, mice may have as many as 150 kb (kilobases) of telomeric DNA per telomere, while the telomeres of *Oxytricha* macronuclear DNA molecules are only 20 bp in length (Kipling and Cooke, Nature 347:400 [1990]; Starling *et al.,* Nucleic Acids Res., 18:6881 [1990]; and Klobutcher *et al.*, Proc. Natl. Acad. Sci., 78:3015 [1981]). Moreover, in most organisms, the amount of telomeric DNA fluctuates. For example, the amount of telomeric DNA at individual yeast telomeres in a wild-type strain may range from approximately 200 to 400 bp, with this amount of DNA increasing and decreasing stoichastically (Shampay and Blackburn, Proc. Natl. Acad. Sci., 85:534 [1988]). Heterogeneity and spontaneous changes in telomere length may reflect a complex balance between the processes involved in degradation and lengthening of telomeric tracts. In addition, genetic, nutritional and other factors may cause increases or decreases in telomeric length (Lustig and Petes, Natl. Acad. Sci., 83:1398 [1986]; and Sandell *et al.,* Cell 91:12061 [1994]). The inherent heterogeneity of virtually all telomeric DNAs suggests that telomeres are not maintained via conventional replicative processes.

In addition to the telomeres themselves, the regions located adjacent to telomeres have been studied. For example, in most organisms, the sub-telomeric regions immediately internal to the simple repeats consist of middle repetitive sequences, designated as telomere-associated ("TA") DNA. These regions bear some similarity with the transposon telomeres of *Drosophila*. In *Saccharomyces*, two classes of TA elements, designated as "X" and "Y,"' have been described (Chan and Tye, Cell 33:563 [1983]). These elements may be found alone or in combination on most or all telomeres.

Telomeric Structural Proteins

Various structural proteins that interact with telomeric DNA have been described which are distinct from the protein components of the telomerase enzyme. Such structural proteins comprise the "telosome" of *Saccharomyces* chromosomes (Wright et al., Genes Develop., 6:197 [1992]) and of ciliate macronuclear DNA molecules (Gottschling and Cech, Cell 38:501 [1984]; and Blackburn and Chiou, Proc. Natl. Acad. Sci., 78:2263 [1981]). The telosome is a non-nucleosomal, but discrete chromatin structure that encompasses the entire terminal array of telomeric repeats. In *Saccharomyces*, the DNA adjacent to the telosome is packaged into nucleosomes. However, these nucleosomes are reported to differ from those in most other regions of the yeast genome, as they have features that are characteristic of transcriptionally inactive chromatin (Wright *et al.*, Genes Develop., 6:197 [1992]; and Braunstein et al., Genes Develop., 7:592 [1993]). In mammals, most of the simple repeated telomeric DNA is packaged in closely spaced nucleosomes (Makarov et al., Cell 73:775 [1993]; and Tommerup et al., Mol. Cell. Biol., 14:5777 [1994]). However, the telomeric repeats located at the very ends of the human chromosomes are found in a telosome-like structure.

Telomere Replication

Complete replication of the ends of linear eukaryotic chromosomes presents special problems for conventional methods of DNA replication. For example, conventional DNA polymerases cannot begin DNA synthesis *de novo*, rather, they require RNA primers which are later removed during replication. In the case of telomeres, removal of the RNA primer from the lagging-strand end would necessarily leave a 5'-terminal gap, resulting in the loss of sequence if the parental telomere was blunt-ended (Watson, Nature New Biol., 239:197 [1972]; Olovnikov, J. Theor. Biol., 41:181 [1973]). However, the described telomeres have 3' overhangs (Klobutcher *et al.*, Proc. Natl. Acad. Sci., 58:3015 [1981]; Henderson and

Blackburn, Mol. Cell. Biol., 9:345 [1989]; and Wellinger et al., Cell 72:51 [1993]). For these molecules, it is possible that removal of the lagging-strand 5'-terminal RNA primer could regenerate the 3' overhang without loss of sequence on this side of the molecule. However, loss of sequence information on the leading-strand end would occur, because of the lack of a complementary strand to act as template in the synthesis of a 3' overhang (Zahler and Prescott, Nucleic Acids Res., 16:6953 [1988]; Lingner et al., Science 269:1533 [1995]).

Nonetheless, complete replication of the chromosomes must occur. While conventional DNA polymerases cannot accurately reproduce chromosomal DNA ends, specialized factors exist to ensure their complete replication. Telomerase is a key component in this process. Telomerase is a ribonucleoprotein (RNP) particle and polymerase that uses a portion of its internal RNA moiety as a template for telomere repeat DNA synthesis (Yu *et al.,* Nature 344:126 [1990]; Singer and Gottschling, Science 266:404 [1994]; Autexier and Greider, Genes Develop., 8:563 [1994]; Gilley *et al.,* Genes Develop., 9:2214 [1995]; McEachern and Blackburn, Nature 367:403 [1995]; Blackburn, Ann. Rev. Biochem., 61:113 [1992];. Greider, Ann. Rev. Biochem., 65:337 [1996]). The activity of this enzyme depends upon both its RNA and protein components to circumvent the problems presented by end replication by using RNA (*i.e.,* as opposed to DNA) to template the synthesis of telomeric DNA. Telomerases extend the G strand of telomeric DNA. A combination of factors, including telomerase processivity, frequency of action at individual telomeres, and the rate of degradation of telomeric DNA, contribute to the size of the telomeres (*i.e.,* whether they are lengthened, shortened, or maintained at a certain size). *In vitro*, telomerases may be extremely processive, with the *Tetrahymena* telomerase adding an average of approximately 500 bases to the G strand primer before dissociation of the enzyme (Greider, Mol. Cell. Biol., 114572 [1991]).

Importantly, telomere replication is regulated both by developmental and cell cycle factors. It has been hypothesized that aspects of telomere replication may act as signals in the cell cycle. For example, certain DNA structures or DNA-protein complex formations may act as a checkpoint to indicate that chromosomal replication has been completed (*See e.g.,* Wellinger *et al.,* Mol. Cell. Biol., 13:4057 [1993]). In addition, it has been observed that in humans, telomerase activity is not detectable in most somatic tissues, although it is detected in many tumors (Wellinger, *supra*). This telomere length may serve as a mitotic clock, which serves to limit the replication potential of cells *in vivo* and/or *in vitro*. What remains needed in the art is a method to study the role of telomeres and their replication in normal as well as abnormal cells (*i.e.,* cancerous cells). An understanding of telomerase and its function is

needed in order to develop means for use of telomerase as a target for cancer therapy or anti-aging processes.

SUMMARY OF THE INVENTION

The present invention provides compositions and methods for purification and use of telomerase. In particular, the present invention is directed to telomerase and co-purifying polypeptides obtained from *Euplotes aediculatus*, as well as other organisms (*e.g.*, *Schizosaccharomyces*, *Tetrahymena*, and humans). The present invention also provides methods useful for the detection and identification of telomerase homologs in other species and genera of organisms.

The present invention provides heretofore unknown telomerase subunit proteins of *E. aediculatus* of approximately 123 kDa and 43 kDa, as measured on SDS-PAGE. In particular, the present invention provides substantially purified 123 kDa and 43 kDa telomerase protein subunits.

One aspect of the invention features isolated and substantially purified polynucleotides which encode telomerase subunits (*i.e.*, the 123 kDa and 43 kDa protein subunits). In a particular aspect, the polynucleotide is the nucleotide sequence of SEQ ID NO:1, or variants thereof. In an alternative embodiment, the present invention provides fragments of the isolated (*i.e.*, substantially purified) polynucleotide encoding the telomerase 123 kDa subunit of at least 10 amino acid residues in length. The invention further contemplates fragments of this polynucleotide sequence (*i.e.*, SEQ ID NO:1) that are at least 6 nucleotides, at least 25 nucleotides, at least 30 nucleotides, at least 100 nucleotides, at least 250 nucleotides, and at least 500 nucleotides in length. In addition, the invention features polynucleotide sequences that hybridize under stringent conditions to SEQ ID NO:1, or fragments thereof. The present invention further contemplates a polynucleotide sequence comprising the complement of the nucleic acid of SEQ ID NO:1, or variants thereof.

The present invention also provides the polynucleotide with the sequence of SEQ ID NO:3. In particular, the present invention provides the polynucleotide sequence comprising at least a portion of the nucleic acid sequence of SEQ ID NO:3, or variants, thereof. In one embodiment, the present invention provides fragments of the isolated (*i.e.*, substantially purified) polynucleotide encoding the telomerase 43 kDa subunit of at least 10 amino acid residues in length. The invention also provides an isolated polynucleotide sequence encoding the polypeptide of SEQ ID NOS:4-6, or variants thereof. The invention further contemplates fragments of this polynucleotide sequence (*i.e.*, SEQ ID NO:3) that are at least 5 nucleotides,

at least 20 nucleotides, at least 100 nucleotides, at least 250 nucleotides, and at least 500 nucleotides in length. In addition, the invention features polynucleotide sequences that hybridize under stringent conditions to SEQ ID NO:3, or fragments thereof. The present invention further contemplates a polynucleotide sequence comprising the complement of the nucleic acid of SEQ ID NO:3, or variants thereof.

The present invention provides a substantially purified polypeptide comprising at least a portion of the amino acid sequence of SEQ ID NO:2, or variants thereof. In one embodiment, the portion of the polypeptide sequence comprises fragments of SEQ ID NO:2, having a length greater than 10 amino acids. However, the invention also contemplates polypeptide sequences of various lengths, the sequences of which are included within SEQ ID NO:2, ranging from 5-500 amino acids. The present invention also provides an isolated polynucleotide sequence encoding the polypeptide of SEQ ID NO:2, or variants, thereof.

The present invention provides a substantially purified polypeptide comprising at least a portion of the amino acid sequence selected from the group consisting of SEQ ID NO:4-6, or variants thereof. In one embodiment, the portion of the polypeptide comprises fragments of SEQ ID NO:4, having a length greater than 10 amino acids. In an alternative embodiment, the portion of the polypeptide comprises fragments of SEQ ID NO:5, having a length greater than 10 amino acids. In yet another alternative embodiment, the portion of the polypeptide comprises fragments of SEQ ID NO:6, having a length greater than 10 amino acids. The present invention also contemplates polypeptide sequences of various lengths, the sequences of which are included within SEQ ID NOS:4, 5, and/or 6, ranging from 5 to 500 amino acids.

The present invention also provides a telomerase complex comprised of at least one purified 123 kDa telomerase protein subunit, at least one a purified 43 kDa telomerase protein subunit, and purified RNA. In a preferred embodiment, the telomerase complex comprises one purified 123 kDa telomerase protein subunit, one purified 43 kDa telomerase protein subunit, and purified telomerase RNA. In one preferred embodiment, the telomerase complex comprises an 123 kDa and/or telomerase protein subunit obtained from *Euplotes aediculatus*. It is contemplated that the 123 kDa telomerase protein subunit of the telomerase complex be encoded by SEQ ID NO:1. It is also contemplated that the 123 kDa telomerase protein subunit of the telomerase complex be comprised of SEQ ID NO:2. It is also contemplated that the 43 kDa telomerase protein subunit of the telomerase complex be obtained from *Euplotes aediculatus*. It is further contemplated that the 43 kDa telomerase subunit of the telomerase complex be encoded by SEQ ID NO:3. It is also contemplated that the 43 kDa telomerase protein subunit of the telomerase complex be comprised of the amino

acid sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:6. It is contemplated that the purified RNA of the telomerase complex be comprised of the RNA encoded by such sequences as those disclosed by Linger *et al.,* (Lingner *et al.*, Genes Develop., 8:1985 [1994]). In a preferred embodiment, the telomerase complex is capable of replicating telomeric DNA.

The present invention also provides methods for identifying telomerase protein subunits in eukaryotic organisms other than *E. aediculatus*. These methods are comprised of multiple steps. The first step is the synthesis of at least one probe or primer oligonucleotide that encodes at least a portion of the amino acid sequence of SEQ ID NOS:2, 4, 5, or 6. In the alternative, the synthesized probe or primer oligonucleotides are complementary to at least a portion of the amino acid sequence of SEQ ID NO:2, 4, 5, or 6. The next step comprises exposing at least one of the probe or primer oligonucleotide(s) to nucleic acid comprising the genome or, in the alternative, the expressed portion of the genome of the other organism (*i.e.,* the non-*E. aediculatus* organism), under conditions suitable for the formation of nucleic acid hybrids. Next, the hybrids are identified with or without amplification, using a DNA polymerase (*e.g., Taq*, or any other suitable polymerase known in the art). Finally, the sequence of the hybrids are determined using methods known in the art, and the sequences of the derived amino acid sequences analyzed for their similarity to SEQ ID NOS:2, 4, 5, or 6.

The present invention also provides methods for identifying nucleic acid sequences encoding telomerase protein subunits in eukaryotic organisms comprising the steps of: providing a sample suspected of containing nucleic acid encoding an eukaryotic telomerase protein subunit; at least one oligonucleotide primer complementary to the nucleic acid sequence encoding at least a region of an *Euplotes aediculatus* telomerase protein subunit; and iii) a polymerase; exposing the sample to the at least one oligonucleotide primer and the polymerase under conditions such that the nucleic acid encoding the eukaryotic telomerase protein subunit is amplified; determining the sequence of the eukaryotic telomerase protein subunit; and comparing the sequence of the eukaryotic telomerase protein subunit and the *Euplotes aediculatus* telomerase protein subunit. In one preferred embodiment, the *Euplotes aediculatus* telomerase subunit comprises at least a portion of SEQ ID NO:1. In an alternative preferred embodiment, the *Euplotes aediculatus* telomerase subunit comprises at least a portion of SEQ ID NO:3.

Thus, the present invention also provides methods for identification of telomerase protein subunits in eukaryotic organisms other than *E. aediculatus*. In addition, the present

invention provides methods for comparisons between the amino acid sequences of SEQ ID NOS:2, 4, 5, or 6, and the amino acid sequences derived from gene sequences of other organisms or obtained by direct amino acid sequence analysis of protein. The amino acid sequences shown to have the greatest degree of identity (*i.e.*, homology) to SEQ ID NOS:2, 4, 5, or 6, may then selected for further testing. Sequences of particular importance are those that share identity with the reverse transcriptase motif of the *Euplotes* sequence. Once identified, the proteins with the sequences showing the greatest degree of identity may be tested for their role in telomerase activity by genetic or biochemical methods, including the methods set forth in the Examples below.

The present invention also provides methods for purification of telomerase comprising the steps of providing a sample containing telomerase, an affinity oligonucleotide, a displacement oligonucleotide; exposing the sample to the affinity oligonucleotide under conditions wherein the affinity oligonucleotide binds to the telomerase to form a telomerase-oligonucleotide complex; and exposing the oligonucleotide-telomerase complex to the displacement oligonucleotide under conditions such that the telomerase is released from the template. In a preferred embodiment, the method comprises the further step of eluting the telomerase. In another preferred embodiment, the affinity oligonucleotide comprises an antisense portion and a biotin residue. It is contemplated that during the exposing step, the biotin residue of the affinity oligonucleotide binds to an avidin bead and the antisense portion binds to the telomerase. It is also contemplated that during the exposing step, the displacement oligonucleotide binds to the affinity oligonucleotide.

The present invention further provides substantially purified polypeptides comprising the amino acid sequence comprising SEQ ID NOS: 63, 64, 65, 67, 68, and 69. In another embodiment, the present invention also provides purified, isolated polynucleotide sequences encoding the polypeptides comprising the amino acid sequences of SEQ ID NOS: 63, 64, 65, 67, 68, and 69. The present invention contemplates portions or fragments of SEQ ID NOS: 63, 64, 65, 67, 68, and 69, of various lengths. In one embodiment, the portion of polypeptide comprises fragments of lengths greater than 10 amino acids. However, the present invention also contemplates polypeptide sequences of various lengths, the sequences of which are included within SEQ ID NOS: 63, 64, 65, 67, 68, and 69, ranging from 5 to 500 amino acids (as appropriate, based on the length of SEQ ID NOS: 63, 64, 65, 67, 68, and 69).

The present invention also provides nucleic acid sequences comprising SEQ ID NOS:55, 62, 66, and 68, or variants thereof. The present invention further provides fragments of the isolated polynucleotide sequences that are at least 6 nucleotides, at least 25

nucleotides, at least 30 nucleotides, at least 50 nucleotides, at least 100 nucleotides, at least 250 nucleotides, and at least 500 nucleotides in length (as appropriate for the length of the sequence of SEQ ID NOS:55, 62, 66, and 68, or variants thereof).

In particularly preferred embodiments, the polynucleotide hybridizes specifically to telomerase sequences, wherein the telomerase sequences are selected from the group consisting of human, *Euplotes aediculatus*, *Oxytricha, Schizosaccharomyces*, and *Saccharomyces* telomerase sequences. In other preferred embodiments, the present invention provides polynucleotide sequences comprising the complement of nucleic acid sequences selected from the group consisting of SEQ ID NOS:55, 62, 66, and 68, or variants thereof. In yet other preferred embodiments, the present invention provides polynucleic acid sequences that hybridize under stringent conditions to at least one nucleic acid sequence selected from the group consisting of SEQ ID NO:55, 62, 66, and 68. In a further embodiment, the polynucleotide sequence comprises a purified, synthetic nucleotide sequence having a length of about ten to thirty nucleotides.

In alternative preferred embodiments, the present invention provides polynucleotide sequences corresponding to the human telomerase, including SEQ ID NOS:173 and 224, and their complementary sequences. The invention further contemplates fragments of these polynucleotide sequence (*i.e.*, SEQ ID NOS: 173 and 224) that are at least 5 nucleotides, at least 20 nucleotides, at least 100 nucleotides, at least 250 nucleotides, and at least 500 nucleotides in length. The invention further contemplates fragments of the complements of these polynucleotide sequences (*i.e.*, SEQ ID NOS: 173 and 224) that are at least 5 nucleotides, at least 20 nucleotides, at least 100 nucleotides, at least 250 nucleotides, and at least 500 nucleotides in length. In addition, the invention features polynucleotide sequences that hybridize under stringent conditions to SEQ ID NOS: 173 and 224, and/or fragments, and/or the complementary sequences thereof. The present invention further contemplates a polynucleotide sequence comprising the complement of the nucleic acids of SEQ ID NOS: 173 and 224, or variants thereof. In a further embodiment, the polynucleotide sequence comprises a purified, synthetic nucleotide sequence corresponding to a fragment of SEQ ID NOS: 173 and 224, having a length of about ten to thirty nucleotides. The present invention further provides plasmid pGRN121 (ATCC accession ##20916), and the lambda clone 25-1.1 (ATCC accession #____).

The present invention further provides substantially purified polypeptides comprising the amino acid sequence comprising SEQ ID NOS:174-223 and 225. In another embodiment, the present invention also provides purified, isolated polynucleotide sequences

encoding the polypeptides comprising the amino acid sequences of SEQ ID NOS: 174-223 and 225. The present invention contemplates portions or fragments of SEQ ID NOS: 174-223 and 225, of various lengths. In one embodiment, the portion of polypeptide comprises fragments of lengths greater than 10 amino acids. However, the present invention also contemplates polypeptide sequences of various lengths, the sequences of which are included within SEQ ID NOS: 174-223 and 225, ranging from 5 to 1100 amino acids (as appropriate, based on the length of SEQ ID NOS: 174-223 and 225).

The present invention also provides methods for detecting the presence of nucleotide sequences encoding at least a portion of human telomerase in a biological sample, comprising the steps of, providing: a biological sample suspected of containing nucleic acid corresponding to the nucleotide sequence set forth in SEQ ID NO:62; the nucleotide of SEQ ID NO:62 or fragment(s) thereof; combining the biological sample with the nucleotide under conditions such that a hybridization complex is formed between the nucleic acid and the nucleotide; and detecting the hybridization complex.

In one embodiment of the method the nucleic acid corresponding to the nucleotide sequence of SEQ ID NO:62, is ribonucleic acid, while in an alternative embodiment, the nucleotide sequence is deoxyribonucleic acid. In yet another embodiment of the method the detected hybridization complex correlates with expression of the polynucleotide of SEQ ID NO:62, in the biological sample. In yet another embodiment of the method, detection of the hybridization complex comprises conditions that permit the detection of alterations in the polynucleotide of SEQ ID NO:62 in the biological sample.

The present invention also provides antisense molecules comprising the nucleic acid sequence complementary to at least a portion of the polynucleotide of SEQ ID NO:55, 62, 66, 67, and 68. In an alternatively preferred embodiment, the present invention also provides pharmaceutical compositions comprising antisense molecules of SEQ ID NOS:55, 62, 67, and 68, and a pharmaceutically acceptable excipient and/or other compound (*e.g.,* adjuvant).

In yet another embodiment, the present invention provides polynucleotide sequences contained on recombinant expression vectors. In one embodiment, the expression vector containing the polynucleotide sequence is contained within a host cell.

The present invention also provides methods for producing polypeptides comprising the amino acid sequence of SEQ ID NOS:61, 63, 65, 67, or 68, the method comprising the steps of: culturing a host cell under conditions suitable for the expression of the polypeptide; and recovering the polypeptide from the host cell culture.

The present invention also provides purified antibodies that binds specifically to a polypeptide comprising at least a portion of the amino acid sequence of SEQ ID NOS:55, 63, 64, 65, 67, and/or 69. In one embodiment, the present invention provides a pharmaceutical composition comprising at least one antibody, and a pharmaceutically acceptable excipient.

The present invention further provides methods for the detection of human telomerase in a biological sample comprising the steps of: providing a biological sample suspected of expressing human telomerase protein; and at least one antibody that binds specifically to at least a portion of the amino acid sequence of SEQ ID NOS:55, 61, 63, 64, 65, 67, and/or 69; combining the biological sample and antibody(ies) under conditions such that an antibody:protein complex is formed; and detecting the complex wherein the presence of the complex correlates with the expression of the protein in the biological sample.

The present invention further provides substantially purified peptides comprising the amino acid sequence selected from the group consisting of SEQ ID NOS:71, 73, 75, 77, 79, 82, 83, 85, and 101. In an alternative embodiment, the present invention provides purified, isolated polynucleotide sequences encoding the polypeptide corresponding to these sequences. In preferred embodiments, the polynucleotide hybridizes specifically to telomerase sequences, wherein the telomerase sequences are selected from the group consisting of human, *Euplotes aediculatus*, *Oxytricha, Schizosaccharomyces*, *Saccharomyces* and *Tetrahymena* telomerase sequences. In yet another embodiment, the polynucleotide sequence comprises the complement of a nucleic acid sequence selected from the group consisting of SEQ ID NOS:70, 72, 74, 76, 78, 80, 81, 100, 173, 224, and variants thereof. In a further embodiment, the polynucleotide sequence that hybridizes under stringent conditions to a nucleic acid sequence selected from the group consisting of SEQ ID NOS:66, 68, 80, and 81. In yet another embodiment, the polynucleotide sequence is selected from the group consisting of SEQ ID NOS:70, 72, 74, 76, 78, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 173, and 224. In an alternative embodiment, the nucleotide sequence comprises a purified, synthetic nucleotide sequence having a length of about ten to fifty nucleotides.

The present invention also provides methods for detecting the presence of nucleotide sequences encoding at least a portion of human telomerase in a biological sample, comprising the steps of, providing: a biological sample suspected of containing nucleic acid corresponding to the nucleotide sequence of SEQ ID NO:100, and/or SEQ ID NO:173, and/or SEQ ID NO:224; the nucleotide of SEQ ID NO:100, and/or SEQ ID NO:173, and/or SEQ ID NO:224, or fragment(s) thereof; combining the biological sample with the nucleotide under

conditions such that a hybridization complex is formed between the nucleic acid and the nucleotide; and detecting the hybridization complex.

In one embodiment of the method the nucleic acid corresponding to the nucleotide sequence of SEQ ID NO:100, and/or SEQ ID NO:173, and/or SEQ ID NO:224, is ribonucleic acid, while in an alternative embodiment, the nucleotide sequence is deoxyribonucleic acid. In yet another embodiment of the method the detected hybridization complex correlates with expression of the polynucleotide of SEQ ID NO:100, and/or SEQ ID NO:173, and/or SEQ ID NO:224, in the biological sample. In yet another embodiment of the method, detection of the hybridization complex comprises conditions that permit the detection of alterations in the polynucleotide of SEQ ID NO:100 and/or SEQ ID NO:173, and/or SEQ ID NO:224, in the biological sample.

The present invention also provides antisense molecules comprising the nucleic acid sequence complementary to at least a portion of the polynucleotide of SEQ ID NOS:82, 100, 173 and 224. In an alternatively preferred embodiment, the present invention also provides pharmaceutical compositions comprising antisense molecules of SEQ ID NOS:82, 100, 173, 224, and a pharmaceutically acceptable excipient and/or other compound (*e.g.,* adjuvant).

In yet another embodiment, the present invention provides polynucleotide sequences contained on recombinant expression vectors. In one embodiment, the expression vector containing the polynucleotide sequence is contained within a host cell.

The present invention also provides methods for producing polypeptides comprising the amino acid sequence of SEQ ID NOS:82, 83, 84, 85, 86, 101, 174-223 and/or 225, the method comprising the steps of: culturing a host cell under conditions suitable for the expression of the polypeptide; and recovering the polypeptide from the host cell culture.

The present invention also provides purified antibodies that binds specifically to a polypeptide comprising at least a portion of the amino acid sequence of SEQ ID NOS:71, 73, 75, 77, 79, 82, 83, 84, 85, 101, 174-223 and/or 225. In one embodiment, the present invention provides a pharmaceutical composition comprising at least one antibody, and a pharmaceutically acceptable excipient.

The present invention further provides methods for the detection of human telomerase in a biological sample comprising the steps of: providing a biological sample suspected of expressing human telomerase protein; and at least one antibody that binds specifically to at least a portion of the amino acid sequence of SEQ ID NOS:71, 73, 75, 77, 79, 82, 83, 84, 85, 87, 101, 174-223 and/or 225, combining the biological sample and antibody(ies) under conditions such that an antibody:protein complex is formed; and detecting the complex

wherein the presence of the complex correlates with the expression of the protein in the biological sample.

DESCRIPTION OF THE FIGURES

Figure 1 is a schematic diagram of the affinity purification of telomerase showing the binding and displacement elution steps.

Figure 2 is a photograph of a Northern blot of telomerase preparations obtained during the purification protocol.

Figure 3 shows telomerase activity through the purification protocol.

Figure 4 is a photograph of a SDS-PAGE gel, showing the presence of an approximately 123 kDa polypeptide and an approximately 43 kDa doublet.

Figure 5 is a graph showing the sedimentation coefficient of telomerase.

Figure 6 is a photograph of a polyacrylamide/urea gel with 36% formamide.

Figure 7 shows the putative alignments of telomerase RNA template, with SEQ ID NOS:43 and 44 in Panel A, and SEQ ID NOS:45 and 46 in Panel B.

Figure 8 is a photograph of lanes 25-30 of the gel shown in Figure 6, shown at a lighter exposure level.

Figure 9 shows the DNA sequence of the gene encoding the 123 kDa telomerase protein subunit (SEQ ID NO:1).

Figure 10 shows the amino acid sequence of the 123 kDa telomerase protein subunit (SEQ ID NO:2).

Figure 11 shows the DNA sequence of the gene encoding the 43 kDa telomerase protein subunit (SEQ ID NO:3).

Figure 12 shows the DNA sequence, as well as the amino acid sequences of all three open reading frames of the 43 kDa telomerase protein subunit (SEQ ID NOS:4-6).

Figure 13 shows a sequence comparison between the 123 kDa telomerase protein subunit of *E. aediculatus* (SEQ ID NO:2) and the 80 kDa polypeptide subunit of *T. thermophila* (SEQ ID NO:52).

Figure 14 shows a sequence comparison between the 123 kDa telomerase protein subunit of *E.aediculatus* (SEQ ID NO:2) and the 95 kDa telomerase polypeptide of *T. thermophila* (SEQ ID NO:54).

Figure 15 shows the best-fit alignment between a portion of the "La-domain" of the 43 kDa telomerase protein subunit of *E. aediculatus* (SEQ ID NO:9) and a portion of the 95 kDa polypeptide subunit of *T. thermophila* (SEQ ID NO:10).

Figure 16 shows the best-fit alignment between a portion of the "La-domain" of the 43 kDa telomerase protein subunit of *E. aediculatus* (SEQ ID NO:11) and a portion of the 80 kDa polypeptide subunit of *T. thermophila* (SEQ ID NO:12).

Figure 17 shows the alignment and motifs of the polymerase domain of the 123 kDa telomerase protein subunit of *E. aediculatus* (SEQ ID NOS:13 and 18) and the polymerase domains of various reverse transcriptases (SEQ ID NOS:14-17, and 19-22).

Figure 18 shows the alignment of a domain of the 43 kDa telomerase protein subunit (SEQ ID NO:23) with various La proteins (SEQ ID NOS:24-27).

Figure 19 shows the nucleotide sequence encoding the *T. thermophila* 80 kDa protein subunit (SEQ ID NO:51).

Figure 20 shows the amino acid sequence of the *T. thermophila* 80 kDa protein subunit (SEQ ID NO:52).

Figure 21 shows the nucleotide sequence encoding the *T. thermophila* 95 kDa protein subunit (SEQ ID NO:53).

Figure 22 shows the amino acid sequence of the *T, thermophila* 95 kDa protein subunit (SEQ ID NO:54).

Figure 23 shows the amino acid sequence of L8543.12 ("Est2p") (SEQ ID NO:55).

Figure 24 shows the alignment of the *Oxytricha* PCR product (SEQ ID NO:58) with the *Euplotes* sequence (SEQ ID NO:59).

Figure 25 shows the alignment of the human telomere amino acid motifs (SEQ ID NO:67), with portions of the tez1 sequence (SEQ ID NO:63), Est2p (SEQ ID NO:64), and the *Euplotes* p123 (SEQ ID NO:65).

Figure 26 shows the DNA sequence of Est2 (SEQ ID NO:66).

Figure 27 shows the amino acid sequence of a cDNA clone (SEQ ID NO:67) encoding human telomerase peptide motifs.

Figure 28 shows the DNA sequence of a cDNA clone (SEQ ID NO:62) encoding human telomerase peptide motifs.

Figure 29 shows the amino acid sequence of *tez1* (SEQ ID NO:69).

Figure 30 shows the DNA sequence of *tez1* (SEQ ID NO:68).

Figure 31 shows the alignment of EST2p (SEQ ID NO:83), *Euplotes* (SEQ ID NO:84), and *Tetrahymena* (SEQ ID NO:85) sequences, as well as consensus sequence.

Figure 32 shows the sequences of peptides useful for production of antibodies.

Figure 33 is a schematic summary of the *tez1*$^{+}$sequencing experiments.

Figure 34 (SEQ ID NOS:118-121)shows two degenerate primers used in PCR to identify the *S. pombe* homolog of the *E. aediculatus* p123 sequences.

Figure 35 (SEQ ID NOS:119 and 121) shows the four major bands produced in PCR using the degenerate primers.

Figure 36 (SEQ ID NOS:58, 118, 121-130) shows the alignment of the M2 PCR product with *E. aediculatus* p123, *S. cerevisiae*, and *Oxytricha* telomerase protein sequences.

Figure 37 (SEQ ID NOS:131 and 132) is a schematic showing the 3' RT PCR strategy.

Figure 38 shows the libraries and the results of screening libraries for *S. pombe* telomerase protein sequences.

Figure 39 shows the results obtained with the *Hin*dIII-digested positive genomic clones containing *S. pombe* telomerase sequence.

Figure 40 is a schematic showing the 5' RT PCR strategy.

Figure 41 (SEQ ID NOS:133-147) shows the alignment of RT domains from telomerase catalytic subunits.

Figure 42 (SEQ ID NOS:2, 55 and 69) shows the alignment of three telomerase sequences.

Figure 43 shows the disruption strategy used with the telomerase genes in *S. pombe*.

Figure 44 shows the experimental results confirming disruption of *tez1*.

Figure 45 shows the progressive shortening of telomeres in *S. pombe* due to *tez1* disruption.

Figure 46 shows the DNA (SEQ ID NO:68) and amino acid (SEQ ID NO:69) sequence of *tez1*, with the coding regions indicated.

Figure 47 shows the DNA (SEQ ID NO:100) and amino acid (SEQ ID NO:101) of the ORF encoding an approximately 63 kDa telomerase protein or fragment thereof.

Figure 48 (SEQ ID NOS:148-171) shows an alignment of reverse transcriptase motifs from various sources.

Figure 49 provides a restriction and function map of plasmid pGRN121.

Figure 50 provides the results of preliminary nucleic acid sequencing analysis of human telomerase (SEQ ID NO:173).

Figure 51 provides the preliminary nucleic acid (SEQ ID NO:173) and deduced ORF sequences (SEQ ID NOS:174-223) of human telomerase.

Figure 52 provides a refined restriction and function map of plasmid pGRN121.

Figure 53 provides the nucleic acid (SEQ ID NO:224) and deduced ORF sequence (SEQ ID NO:225) of human telomerase.

Figure 54 provides a restriction map of lambda clone 25-1.1 (ATCC accession #_____).

DEFINITIONS

To facilitate understanding the invention, a number of terms are defined below.

As used herein, the term "ciliate" refers to any of the protozoans belonging to the phylum Ciliaphora.

As used herein, the term "eukaryote" refers to organisms distinguishable from "prokaryotes." It is intended that the term encompass all organisms with cells that exhibit the usual characteristics of eukaryotes such as the presence of a true nucleus bounded by a nuclear membrane, within which lie the chromosomes, the presence of membrane-bound organelles, and other characteristics commonly observed in eukaryotic organisms. Thus, the term includes, but is not limited to such organisms as fungi, protozoa, and animals (*e.g.*, humans).

As used herein, the term "polyploid" refers to cells or organisms which contain more than two sets of chromosomes.

As used herein, the term "macronucleus" refers to the larger of the two types of nuclei observed in the ciliates. This structure is also sometimes referred to as the "vegetative" nucleus. Macronuclei contain many copies of each gene and are transcriptionally active.

As used herein, the term "micronucleus" refers to the smaller of the two types of nuclei observed in the ciliates. This structure is sometimes referred to as the "reproductive" nucleus, as it participates in meiosis and autogamy. Micronuclei are diploid and are transcriptionally inactive.

As used herein, the term "ribonucleoprotein" refers to a complex macromolecule containing both RNA and protein.

As used herein, the term "telomerase polypeptide," refers to a polypeptide which is at least a portion of the *Euplotes* telomerase structure. The term encompasses the 123 kDa and 43 kDa polypeptide or protein subunits of the *Euplotes* telomerase. It is also intended that the term encompass variants of these protein subunits. It is further intended to encompass the polypeptides encoded by SEQ ID NOS: 1 and 3. As molecular weight measurements may

vary, depending upon the technique used, it is not intended that the present invention be precisely limited to the 123 kDa or 43 kDa molecular masses of the polypeptides encoded by SEQ ID NOS:1 and 3, as determined by any particular method such as SDS-PAGE.

As used herein, the terms "telomerase" and "telomerase complex" refer to functional telomerase enzymes. It is intended that the terms encompass the complex of proteins and nucleic acids found in telomerases. For example, the terms encompass the 123 kDa and 43 kDa telomerase protein subunits and RNA of *E. aediculatus*.

As used herein, the term "capable of replicating telomeric DNA" refers to functional telomerase enzymes which are capable of performing the function of replicating DNA located in telomeres. It is contemplated that this term encompass the replication of telomeres, as well as sequences and structures that are commonly found located in telomeric regions of chromosomes. For example, "telomeric DNA" includes, but is not limited to the tandem array of repeat sequences found in the telomeres of most organisms.

"Nucleic acid sequence" as used herein refers to an oligonucleotide, nucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded, and represent the sense or antisense strand. Similarly, "amino acid sequence" as used herein refers to peptide or protein sequence. "Peptide nucleic acid" as used herein refers to an oligomeric molecule in which nucleosides are joined by peptide, rather than phosphodiester, linkages. These small molecules, also designated anti-gene agents, stop transcript elongation by binding to their complementary (template) strand of nucleic acid (Nielsen *et al.*, Anticancer Drug Des 8:53-63 [1993]).

A "deletion" is defined as a change in either nucleotide or amino acid sequence in which one or more nucleotides or amino acid residues, respectively, are absent.

An "insertion" or "addition" is that change in a nucleotide or amino acid sequence which has resulted in the addition of one or more nucleotides or amino acid residues, respectively, as compared to, naturally occurring sequences.

A "substitution" results from the replacement of one or more nucleotides or amino acids by different nucleotides or amino acids, respectively.

As used herein, the term "purified" refers to the removal of contaminant(s) from a sample. As used herein, the term "substantially purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated. An "isolated polynucleotide" is therefore a substantially purified polynucleotide.

As used herein, the term "probe" refers to an oligonucleotide (*i.e.*, a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of hybridizing to another oligonucleotide or polynucleotide of interest. Probes are useful in the detection, identification and isolation of particular gene sequences. It is contemplated that any probe used in the present invention will be labelled with any "reporter molecule," so that is detectable in any detection system, including, but not limited to enzyme (*e.g.*, ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is further contemplated that the oligonucleotide of interest (*i.e.*, to be detected) will be labelled with a reporter molecule. It is also contemplated that both the probe and oligonucleotide of interest will be labelled. It is not intended that the present invention be limited to any particular detection system or label.

As used herein, the term "target" refers to the region of nucleic acid bounded by the primers used for polymerase chain reaction. Thus, the "target" is sought to be sorted out from other nucleic acid sequences. A "segment" is defined as a region of nucleic acid within the target sequence.

"Amplification" is defined as the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction (PCR) or other technologies well known in the art (*e.g.*, Dieffenbach and Dveksler, *PCR Primer, a Laboratory Manual*, Cold Spring Harbor Press, Plainview NY [1995]). As used herein, the term "polymerase chain reaction" ("PCR") refers to the method of K.B. Mullis (U.S. Patent Nos. 4,683,195 and 4,683,202, hereby incorporated by reference), which describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing and polymerase extension can be repeated many times (*i.e.*, denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to

each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter "PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified".

As used herein, the term "polymerase" refers to any polymerase suitable for use in the amplification of nucleic acids of interest. It is intended that the term encompass such DNA polymerases as *Taq* DNA polymerase obtained from *Thermus aquaticus*, although other polymerases, both thermostable and thermolabile are also encompassed by this definition.

With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (*e.g.*, hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of ^{32}P-labeled deoxynucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any oligonucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications. Amplified target sequences may be used to obtain segments of DNA (*e.g.*, genes) for insertion into recombinant vectors.

As used herein, the terms "PCR product" and "amplification product" refer to the resultant mixture of compounds after two or more cycles of the PCR steps of denaturation, annealing and extension are complete. These terms encompass the case where there has been amplification of one or more segments of one or more target sequences.

As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes, each of which cut double-stranded DNA at or near a specific nucleotide sequence.

As used herein, the term "recombinant DNA molecule" as used herein refers to a DNA molecule which is comprised of segments of DNA joined together by means of molecular biological techniques.

As used herein, the terms "complementary" or "complementarity" are used in reference to polynucleotides (*i.e.*, a sequence of nucleotides) related by the base-pairing rules. For example, for the sequence "A-G-T," is complementary to the sequence "T-C-A." Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic

acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods which depend upon binding between nucleic acids.

The term "homology" refers to a degree of complementarity. There may be partial homology or complete homology (*i.e.*, identity). A partially complementary sequence is one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (*i.e.*, the hybridization) of a completely homologous to a target under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (*i.e.*, selective) interaction. The absence of non-specific binding may be tested by the use of a second target which lacks even a partial degree of complementarity (*e.g.*, less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

The art knows well that numerous equivalent conditions may be employed to comprise either low or high stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (*e.g.*, the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of either low or high stringency hybridization different from, but equivalent to, the above listed conditions. The term "hybridization" as used herein includes "any process by which a strand of nucleic acid joins with a complementary strand through base pairing" (Coombs, *Dictionary of Biotechnology*, Stockton Press, New York NY [1994].

"Stringency" typically occurs in a range from about T_m-5°C (5°C below the T_m of the probe) to about 20°C to 25°C below T_m. As will be understood by those of skill in the art, a stringent hybridization can be used to identify or detect identical polynucleotide sequences or to identify or detect similar or related polynucleotide sequences.

As used herein, the term "T_m" is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the T_m of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41(\% G + C)$, when a nucleic acid is in aqueous solution at 1 M NaCl (*see e.g.,* Anderson and Young, Quantitative Filter Hybridisation, *in Nucleic Acid Hybridisation* (1985). Other references include more sophisticated computations which take structural as well as sequence characteristics into account for the calculation of T_m.

As used herein the term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bounds between complementary G and C bases and between complementary A and T bases; these hydrogen bonds may be further stabilized by base stacking interactions. The two complementary nucleic acid sequences hydrogen bond in an antiparallel configuration. A hybridization complex may be formed in solution (*e.g.*, C_0t or R_0t analysis) or between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized to a solid support (*e.g.*, a nylon membrane or a nitrocellulose filter as employed in Southern and Northern blotting, dot blotting or a glass slide as employed in *in situ* hybridization, including FISH [fluorescent *in situ* hybridization]).

As used herein, the term "antisense" is used in reference to RNA sequences which are complementary to a specific RNA sequence (*e.g.,* mRNA). Antisense RNA may be produced by any method, including synthesis by splicing the gene(s) of interest in a reverse orientation to a viral promoter which permits the synthesis of a coding strand. Once introduced into a cell, this transcribed strand combines with natural mRNA produced by the cell to form duplexes. These duplexes then block either the further transcription of the mRNA or its translation. In this manner, mutant phenotypes may be generated. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. The designation (-) (*i.e.*, "negative") is sometimes used in reference to the antisense strand, with the designation (+) sometimes used in reference to the sense (*i.e.*, "positive") strand.

As used herein the term "portion" when in reference to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from four amino acid residues to the entire amino acid sequence minus one amino acid. Thus, a

protein "comprising at least a portion of the amino acid sequence of SEQ ID NO:2" encompasses the full-length 123 kDa telomerase protein subunit and fragments thereof.

The term "antigenic determinant" as used herein refers to that portion of an antigen that makes contact with a particular antibody (*i.e.*, an epitope). When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to as antigenic determinants. An antigenic determinant may compete with the intact antigen (*i.e.*, the immunogen used to elicit the immune response) for binding to an antibody.

The terms "specific binding" or specifically binding" when used in reference to the interaction of an antibody and a protein or peptide means that the interaction is dependent upon the presence of a particular structure (*i.e.*, the antigenic determinant or epitope) on the protein; in other words the antibody is recognizing and binding to a specific protein structure rather than to proteins in general. For example, if an antibody is specific for epitope "A", the presence of a protein containing epitope A (or free, unlabelled A) in a reaction containing labelled "A" and the antibody will reduce the amount of labelled A bound to the antibody.

The term "sample" as used herein is used in its broadest sense. A biological sample suspected of containing nucleic acid encoding telomerase subunits may comprise a cell, chromosomes isolated from a cell (*e.g.*, a spread of metaphase chromosomes), genomic DNA (in solution or bound to a solid support such as for Southern blot analysis), RNA (in solution or bound to a solid support such as for Northern blot analysis), cDNA (in solution or bound to a solid support) and the like. A sample suspected of containing a protein may comprise a cell, a portion of a tissue, an extract containing one or more proteins and the like.

The term "correlates with expression of a polynucleotide," as used herein, indicates that the detection of the presence of ribonucleic acid (RNA) complementary to a telomerase sequence by hybridization assays is indicative of the presence of mRNA encoding eukaryotic telomerases, including human telomerases in a sample, and thereby correlates with expression of the telomerase mRNA from the gene encoding the protein.

"Alterations in the polynucleotide" as used herein comprise any alteration in the sequence of polynucleotides encoding telomerases, including deletions, insertions, and point mutations that may be detected using hybridization assays. Included within this definition is the detection of alterations to the genomic DNA sequence which encodes telomerase (*e.g.*, by alterations in pattern of restriction enzyme fragments capable of hybridizing to any sequence such as SEQ ID NOS: 1 or 3 [*e.g.*, RFLP analysis], the inability of a selected fragment of any

sequence to hybridize to a sample of genomic DNA [*e.g.*, using allele-specific oligonucleotide probes], improper or unexpected hybridization, such as hybridization to a locus other than the normal chromosomal locus for the telomere or telomerase genes *e.g.*, using FISH to metaphase chromosomes spreads, etc.]).

A "variant" in regard to amino acid sequences is used to indicate an amino acid sequence that differs by one or more amino acids from another, usually related amino acid. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (*e.g.*, replacement of leucine with isoleucine). More rarely, a variant may have "non-conservative" changes, *e.g.*, replacement of a glycine with a tryptophan. Similar minor variations may also include amino acid deletions or insertions (*i.e.*, additions), or both. Guidance in determining which and how many amino acid residues may be substituted, inserted or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNAStar software. Thus, it is contemplated that this definition will encompass variants of telomerase and/or telomerase protein subunits. For example, the polypeptides encoded by the three open reading frames (ORFs) of the 43 kDa polypeptide gene may be considered to be variants of each other. Such variants can be tested in functional assays, such as telomerase assays to detect the presence of functional telomerase in a sample.

The term "derivative" as used herein refers to the chemical modification of a nucleic acid encoding telomerase structures, such as the 123 kDa or 43 kDa protein subunits of the *E. aediculatus* telomerase, or other telomerase proteins or peptides. Illustrative of such modifications would be replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative would encode a polypeptide which retains essential biological characteristics of naturally-occurring telomerase or its subunits.

The term "biologically active" refers to telomerase molecules or peptides having structural, regulatory, or biochemical functions of a naturally occurring telomerase molecules or peptides. Likewise, "immunologically active," defines the capability of the natural, recombinant, or synthetic telomerase proteins or any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells, and to bind with specific antibodies.

"Affinity purification" as used herein refers to the purification of ribonucleoprotein particles, through the use of an "affinity oligonucleotide" (*i.e.*, an antisense oligonucleotides) to bind the particle, followed by the step of eluting the particle from the oligonucleotide by means of a "displacement oligonucleotide." In the present invention, the displacement oligonucleotide has a greater degree of complementarity with the affinity oligonucleotide,

and therefore produces a more thermodynamically stable duplex than the particle and the affinity oligonucleotide. For example, telomerase may be bound to the affinity oligonucleotide and then eluted by use of a displacement oligonucleotide which binds to the affinity oligonucleotide. In essence, the displacement oligonucleotide displaces the telomerase from the affinity oligonucleotide, allowing the elution of the telomerase. Under sufficiently mild conditions, the method results in the enrichment of functional ribonucleoprotein particles. Thus, the method is useful for the purification of telomerase from a mixture of compounds.

GENERAL DESCRIPTION OF THE INVENTION

The present invention provides purified telomerase preparations and telomerase protein subunits useful for investigations of the activities of telomerases, including potential nuclease activities. In particular, the present invention is directed to the telomerase and co-purifying polypeptides obtained from *Euplotes aediculatus*. This organism, a hypotrichous ciliate, was chosen for use in this invention as it contains an unusually large number of chromosomal ends (Prescott, Microbiol. Rev., 58:233 [1994]), because a very large number of gene-sized DNA molecules are present in its polyploid macronucleus. *Tetrahymena,* a holotrichous ciliate commonly used in previous studies of telomerase and telomeres, is as evolutionarily distant from *Euplotes* as plants are from mammals (Greenwood *et al.,* J. Mol. Evol., 3:163 [1991]).

The homology found between the 123 kDa *E. aediculatus* telomerase subunit and the L8543.12 sequence (*i.e.,* Est2 of *Saccharomyces cerevisiae*; *See,* Lendvay *et al.*, Genetics 144:1399-1412 [1996]), *Schizosaccharomyces*, and human motifs, provides a strong basis for predicting that full human telomerase molecule comprises a protein that is large, basic, and includes such reverse transcriptase motifs. Thus, the compositions and methods of the present invention is useful for the identification of other telomerases, from a wide variety of species. The present invention describes the use of the 123 kDa reverse transcriptase motifs in a method to identify similar motifs in organisms that are distantly related to *Euplotes* (*e.g., Oxytricha*), as well as organisms that are not related to *Euplotes* (*e.g., Saccharomyces, Schizosaccharomyces*, humans, etc.).

The present invention also provides additional methods for the study of the structure and function of distinct forms of telomerase. It is contemplated that the telomerase proteins of the present invention will be useful in diagnostic applications, evolutionary (*e.g.,* phylogenetic) investigations, as well as development of compositions and methods for cancer

therapy or anti-aging regimens. Although the telomerase protein subunits of the present invention themselves have utility, it further contemplated that the polypeptides of the present invention will be useful in conjunction with the RNA moiety of the telomerase enzyme (*i.e.*, a complete telomerase).

It is also contemplated that methods and compositions of this invention will lead to the discovery of additional unique telomerase structures and/or functions. In addition, the present invention provides novel methods for purification of functional telomerase, as well as telomerase proteins. This affinity based method described in Example 3, is an important aspect in the purification of functionally active telomerase. A key advantage of this procedure is the ability to use mild elution conditions, during which proteins that bind non-specifically to the column matrix are not eluted.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed to the nucleic and amino acid sequences of the protein subunits of the *E. aediculatus* telomerase, as well as the nucleic and amino acid sequences of the telomerases from other organisms, including humans. In addition, the present invention is directed to the purification of functional telomerase. As described below the present invention also comprises various forms of telomerase, including recombinant telomerase and telomerase protein subunits, obtained from various organisms.

The 123 kDa and 43 kDa Telomerase Subunit Protein Sequences

The nucleic acid and deduced amino acid sequences of the 123 and 43 kDa protein subunits are shown in Figures 1-6. In accordance with the invention, any nucleic acid sequence which encodes *E. aediculatus* telomerase or its subunits can be used to generate recombinant molecules which express the telomerase or its subunits.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of telomerase subunit protein sequences, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. The invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices, taking into account the use of the codon "UGA" as encoding cysteine in *E. aediculatus*. Other than the exception of the "UGA" codon, these combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence encoding naturally occurring *E. aediculatus* telomerase, and all such variations are to be considered as

being specifically disclosed. For example, the amino acid sequences encoded by each of the three open reading frames of the 43 kDa nucleotide sequence are specifically included (SEQ ID NOS:4-6). It is contemplated that any variant forms of telomerase subunit protein be encompassed by the present invention, as long as the proteins are functional in assays such as those described in the Examples.

Although nucleotide sequences which encode *E. aediculatus* telomerase protein subunits and their variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring sequence under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding *E. aediculatus* telomerase protein subunits or their derivatives possessing a substantially different codon usage, including the "standard" codon usage employed by human and other systems. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic expression host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding telomerase subunits and their derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater or a shorter half-life, than transcripts produced from the naturally occurring sequence.

It is now possible to produce a DNA sequence, or portions thereof, encoding telomerase protein subunits and their derivatives entirely by synthetic chemistry, after which the synthetic gene may be inserted into any of the many available DNA vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding *E. aediculatus* protein subunits or any portion thereof, as well as sequences encoding yeast or human telomerase proteins, subunits, or any portion thereof.

Also included within the scope of the present invention are polynucleotide sequences that are capable of hybridizing to the nucleotide sequence of Figures 9, 11, 12, and 26, under various conditions of stringency. Hybridization conditions are based on the melting temperature (T_m) of the nucleic acid binding complex or probe, as taught in Berger and Kimmel (Berger and Kimmel, *Guide to Molecular Cloning Techniques,* Meth. Enzymol., vol. 152, Academic Press, San Diego CA [1987]) incorporated herein by reference, and may be used at a defined "stringency".

Altered nucleic acid sequences encoding telomerase protein subunits which may be used in accordance with the invention include deletions, insertions or substitutions of

different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent telomerase subunit. The protein may also show deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent telomerase subunit. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological activity of the telomerase subunit is retained. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine, valine; glycine, alanine; asparagine, glutamine; serine, threonine; and phenylalanine, tyrosine.

Methods for DNA sequencing are well known in the art and employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp, Cleveland OH), *Taq* DNA polymerase (Perkin Elmer, Norwalk CT), thermostable T7 polymerase (Amersham, Chicago IL), or combinations of recombinant polymerases and proofreading exonucleases such as the ELONGASE Amplification System marketed by Gibco BRL (Gaithersburg MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown MA) and the ABI 377 DNA sequencers (Perkin Elmer).

Also included within the scope of the present invention are alleles encoding human telomerase proteins and subunits. As used herein, the term "allele" or "allelic sequence" is an alternative form of the nucleic acid sequence encoding human telomerase proteins or subunits. Alleles result from mutations (*i.e,*, changes in the nucleic acid sequence), and generally produce altered mRNAs or polypeptides whose structure and/or function may or may not be altered. An given gene may have no, one or many allelic forms. Common mutational changes that give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of amino acids. Each of these types of changes may occur alone, or in combination with the others, one or more times within a given sequence.

Human Telomerase Motifs

The present invention also provides nucleic and amino acid sequence information for human telomerase motifs. These sequences were first identified in a BLAST search conducted using the *Euplotes* 123 kDa peptide, and a homologous sequence from *Schizosaccharomyces*, designated as "tez1." Figure 25 shows the sequence alignment of the *Euplotes* ("p123"), *Schizosaccharomyces* ("tez1"), Est2p (*i.e.,* the *S. cerevisiae* protein encoded by the *Est2* nucleic acid sequence, and also referred to herein as "L8543.12"), and the human homolog identified in this comparison search. The amino acid sequence of this aligned portion is provided in SEQ ID NO:67 (the cDNA sequence is provided in SEQ ID NO:62), while the portion of tez1 shown in Figure 25 is provided in SEQ ID NO:63. The portion of Est2 shown in this Figure is also provided in SEQ ID NO:64, while the portion of p123 shown is also provided in SEQ ID NO:65.

As shown in Figure 25, there are regions that are highly conserved among these proteins. For example, as shown in this Figure, there are regions of identity in "Motif 0," "Motif 1, "Motif 2," and "Motif 3." The identical amino acids are indicated with an asterisk (*), while the similar amino acid residues are indicated by a circle (●). This indicates that there are regions within the telomerase motifs that are conserved among a wide variety of eukaryotes, ranging from yeast to ciliates, to humans. It is contemplated that additional organisms will likewise contain such conserved regions of sequence.

Figure 27 shows the amino acid sequence of the cDNA clone encoding human telomerase motifs (SEQ ID NO:67), while Figure 28 shows the DNA sequence of the clone. Figure 29 shows the amino acid sequence of tez1 (SEQ ID NO:69), while Figure 30 shows the DNA sequence of tez1 (SEQ ID NO:68). In Figure 30, the introns and other non-coding regions are shown in lower case, while the exons (*i.e.,* coding regions are shown in upper case.

Extending The Polynucleotide Sequence

The polynucleotide sequence encoding telomerase, or telomerase protein subunits, or their functional equivalents, may be extended utilizing partial nucleotide sequence and various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, Gobinda *et al.* (PCR Meth. Applic. 2:318-22 [1993]) describe "restriction-site" polymerase chain reaction (PCR) as a direct method which uses universal primers to retrieve unknown sequence adjacent to a known locus. First, genomic DNA is amplified in the presence of primer to a linker sequence and a primer specific to the

known region. The amplified sequences are subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR can be used to amplify or extend sequences using divergent primers based on a known region (Triglia *et al.*, Nucleic Acids Res 16:8186 [1988]). The primers may be designed using OLIGO® 4.06 Primer Analysis Software (National Biosciences Inc, Plymouth MN [1992]), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72°C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Capture PCR (Lagerstrom *et al.* PCR Methods Applic 1:111-19 [1991]), a method for PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA, may also be used. Capture PCR also requires multiple restriction enzyme digestions and ligations to place an engineered double-stranded sequence into an unknown portion of the DNA molecule before PCR.

Another method which may be used to retrieve unknown sequence is walking PCR (Parker *et al.*, Nucleic Acids Res 19:3055-60 [1991]), a method for targeted gene walking. Alternatively, PCR, nested primers, PromoterFinder™ (Clontech, Palo Alto CA) and PromoterFinder libraries can be used to walk in genomic DNA. This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

Preferred libraries for screening for full length cDNAs are ones that have been size-selected to include larger cDNAs. Also, random primed libraries are preferred in that they will contain more sequences which contain the 5' and upstream regions of genes. A randomly primed library may be particularly useful if an oligo d(T) library does not yield a full-length cDNA. Genomic libraries are useful for extension into the 5' nontranslated regulatory region.

Capillary electrophoresis may be used to analyze either the size or confirm the nucleotide sequence in sequencing or PCR products. Systems for rapid sequencing are available from Perkin Elmer, Beckman Instruments (Fullerton CA), and other companies. Capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled devise camera. Output/light intensity is

converted to electrical signal using appropriate software (*e.g.*, Genotyper™ and Sequence Navigator™ from Perkin Elmer) and the entire process from loading of samples to computer analysis and electronic data display is computer controlled. Capillary electrophoresis is particularly suited to the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample. The reproducible sequencing of up to 350 bp of M13 phage DNA in 30 min has been reported (Ruiz-Martinez *et al.*, Anal Chem 65:2851-8 [1993]).

Expression of the Nucleotide Sequence

In accordance with the present invention, polynucleotide sequences which encode telomerase, telomerase protein subunits, or their functional equivalents, may be used in recombinant DNA molecules that direct the expression of telomerase or telomerase subunits by appropriate host cells.

The nucleotide sequences of the present invention can be engineered in order to alter either or both telomerase subunits for a variety of reasons, including but not limited to, alterations which modify the cloning, processing and/or expression of the gene product. For example, mutations may be introduced using techniques which are well known in the art (*e.g.*, site-directed mutagenesis to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to produce splice variants, etc.).

In an alternate embodiment of the invention, the sequence encoding the telomerase subunit(s) may be synthesized, whole or in part, using chemical methods well known in the art (*See e.g.,* Caruthers *et al.*, Nucleic Acids Res. Symp. Ser., 215-223 [1980]; and Horn *et al.* Nucleic Acids Res. Symp. Ser., 225-232 [1980]). Alternatively, the protein itself could be produced using chemical methods to synthesize a telomerase subunit amino acid sequence, in whole or in part. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, *et al.* Science 269:202 [1995]) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer) in accordance with the instructions provided by the manufacturer.

The newly synthesized peptide can be substantially purified by preparative high performance liquid chromatography (*e.g.*, Creighton, *Proteins, Structures and Molecular Principles*, WH Freeman and Co, New York NY [1983]). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (*e.g.*, the Edman degradation procedure; Creighton, *supra*). Additionally the amino acid sequences of

telomerase subunit proteins, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

Expression Systems

In order to express a biologically active telomerase protein subunit, the nucleotide sequence encoding the subunit or the functional equivalent, is inserted into an appropriate expression vector (*i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence). In order to express a biologically active telomerase enzyme, the nucleotide sequence encoding the telomerase protein subunits are inserted into appropriate expression vectors and the nucleotide sequence encoding the telomerase RNA subunit is inserted into the same or another vector for RNA expression. The protein and RNA subunits are then either expressed in the same cell or expressed separately, and then mixed to achieve a reconstituted telomerase.

Methods which are well known to those skilled in the art can be used to construct expression vectors containing a telomerase protein subunit sequence and appropriate transcriptional or translational controls. These methods include *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination or genetic recombination. Such techniques are described in Sambrook *et al.* (Sambrook *et al., Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY [1989]), and Ausubel *et al.* (Ausubel *et al., Current Protocols in Molecular Biology*, John Wiley & Sons, New York NY [1989]). These same methods may be used to convert the UGA codons, which encode cysteine in *Euplotes*, to the UGU or UGC codon for cysteine recognized by the host expression system.

A variety of expression vector/host systems may be utilized to contain and express a telomerase subunit-encoding sequence. These include but are not limited to microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (*e.g.*, baculovirus); plant cell systems transfected with virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with bacterial expression vectors (*e.g.*, Ti or pBR322 plasmid); or animal cell systems.

The "control elements" or "regulatory sequences" of these systems vary in their strength and specificities and are those non-translated regions of the vector, enhancers,

promoters, and 3' untranslated regions, which interact with host cellular proteins to carry out transcription and translation. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid *lacZ* promoter of the Bluescript® phagemid (Stratagene, La Jolla CA) or pSport1 (Gibco BRL) and ptrp-lac hybrids and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (*e.g.*, heat shock, RUBISCO; and storage protein genes) or from plant viruses (*e.g.*, viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from the mammalian genes or from mammalian viruses are most appropriate. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding telomerase or telomerase protein subunits, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for the telomerase protein or subunit. For example, when large quantities of telomerase protein, subunit, or peptides, are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be desirable. Such vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as Bluescript® (Stratagene), in which the sequence encoding the telomerase or protein subunit may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of β-galactosidase so that a hybrid protein is produced (*e.g.*, pIN vectors; Van Heeke and Schuster, J. Biol. Chem., 264:5503-5509 [1989]) and the like. pGEX vectors (Promega, Madison WI) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems are designed to include heparin, thrombin or factor Xa protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase and PGH may be used. For reviews, see Ausubel *et al.* (*supra*) and Grant *et al.*, Meth. Enzymol., 153:516-544 (1987).

In cases where plant expression vectors are used, the expression of a sequence encoding telomerase or protein subunit, may be driven by any of a number of promoters. For

example, viral promoters such as the 35S and 19S promoters of CaMV (Brisson *et al.*, Nature 310:511-514 [1984]) may be used alone or in combination with the omega leader sequence from TMV (Takamatsu *et al.*, EMBO J., 6:307-311 [1987]). Alternatively, plant promoters such as the small subunit of RUBISCO (Coruzzi *et al.* EMBO J., 3:1671-1680 [1984]; Broglie *et al.,* Science 224:838-843 [1984]) or heat shock promoters (Winter and Sinibaldi Results Probl. Cell Differ., 17:85-105 [1991]) may be used. These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (for reviews of such techniques, see Hobbs or Murry, in *McGraw Hill Yearbook of Science and Technology* McGraw Hill New York NY, pp. 191-196 [1992]; or Weissbach and Weissbach, *Methods for Plant Molecular Biology*, Academic Press, New York NY, pp. 421-463 [1988]).

An alternative expression system which could be used to express telomerase or telomerase protein subunit is an insect system. In one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequence encoding the telomerase sequence of interest may be cloned into a nonessential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of the sequence encoding the telomerase protein or telomerase protein subunit will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses are then used to infect *S. frugiperda* cells or *Trichoplusia* larvae in which the telomerase sequence is expressed (Smith *et al.*, J. Virol., 46:584 [1983]; Engelhard *et al.,* Proc. Natl. Acad. Sci. 91:3224-7 [1994]).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, a sequence encoding telomerase protein or telomerase protein subunit, may be ligated into an adenovirus transcription/ translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a nonessential E1 or E3 region of the viral genome will result in a viable virus capable of expressing in infected host cells (Logan and Shenk, Proc. Natl. Acad. Sci., 81:3655-59 [1984]). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be required for efficient translation of a sequence encoding telomerase protein subunits. These signals include the ATG initiation codon and adjacent sequences. In cases where the sequence encoding a telomerase protein subunit, its initiation codon and upstream sequences are inserted into the most appropriate expression

vector, no additional translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous transcriptional control signals including the ATG initiation codon must be provided. Furthermore, the initiation codon must be in the correct reading frame to ensure transcription of the entire insert. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate to the cell system in use (Scharf *et al.*, Results Probl. Cell Differ., 20:125 [1994]; and Bittner *et al.*, Meth. Enzymol., 153:516 [1987).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be important for correct insertion, folding and/or function. Different host cells such as CHO (ATCC CCL 61 and CRL 9618), HeLa (ATCC CCL 2), MDCK (ATCC CCL 34 and CRL 6253), HEK 293 (ATCC CRL 1573), WI-38 (ATCC CCL 75) (ATCC: American Type Culture Collection, Rockville, MD), etc have specific cellular machinery and characteristic mechanisms for such post-translational activities and may be chosen to ensure the correct modification and processing of the introduced, foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express telomerase or a telomerase subunit protein may be transformed using expression vectors which contain viral origins of replication or endogenous expression elements and a selectable marker gene. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clumps of stably transformed cells can be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler *et al.*, Cell 11:223-32 [1977]) and adenine phosphoribosyltransferase (Lowy *et al.*, Cell 22:817 [1980]) genes which can be employed in tk- or aprt- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, *dhfr* which confers resistance to methotrexate (Wigler *et al.*, Proc. Natl. Acad. Sci., 77:3567

[1980]); *npt*, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin *et al.*, J. Mol. Biol., 150:1 [1981]) and *als* or *pat*, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *In McGraw Hill Yearbook of Science and Technology*, McGraw Hill, New York NY, pp 191-196, [1992]). Additional selectable genes have been described, for example, *trpB*, which allows cells to utilize indole in place of tryptophan, or *hisD*, which allows cells to utilize histinol in place of histidine (Hartman and Mulligan, Proc. Natl. Acad. Sci., 85:8047 [1988]). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, β-glucuronidase and its substrate, GUS, and luciferase and its substrate, luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes *et al.*, Meth. Mol. Biol., 55:121 [1995]).

Identification of Transformants Containing the Polynucleotide Sequence

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression should be confirmed. For example, if the sequence encoding a telomerase protein subunit is inserted within a marker gene sequence, recombinant cells containing the sequence encoding the telomerase protein subunit can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with the sequence encoding telomerase protein subunit under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem sequence as well.

Alternatively, host cells which contain the coding sequence for telomerase or a telomerase protein subunit and express the telomerase or protein subunit be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridization and protein bioassay or immunoassay techniques which include membrane, solution, or chip-based technologies for the detection and/or quantification of the nucleic acid or protein.

The presence of the polynucleotide sequence encoding telomerase protein subunits can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes, portions, or fragments of the sequence encoding the subunit. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the nucleic acid sequence to detect transformants containing DNA or RNA encoding the telomerase subunit. As used herein "oligonucleotides" or "oligomers" refer to a nucleic acid sequence of

approximately 10 nucleotides or greater and as many as approximately 100 nucleotides, preferably between 15 to 30 nucleotides, and more preferably between 20-25 nucleotides which can be used as a probe or amplimer. A variety of protocols for detecting and measuring the expression of proteins (*e.g.*, telomerase or a telomerase protein subunits) using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). These and other assays are described, among other places, in Hampton *et al.*, *Serological Methods a Laboratory Manual*, APS Press, St Paul MN [1990]) and Maddox *et al.*, J. Exp. Med., 158:1211 [1983]).

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting related sequences include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, a telomerase protein subunit sequence, or any portion of it, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labeled nucleotides.

A number of companies such as Pharmacia Biotech (Piscataway NJ), Promega (Madison WI), and US Biochemical Corp (Cleveland OH) supply commercial kits and protocols for these procedures. Suitable reporter molecules or labels include those radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like. Patents teaching the use of such labels include U.S. Patents 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241, herein incorporated by reference. Also, recombinant immunoglobulins may be produced as shown in U.S. Patent No. 4,816,567 incorporated herein by reference.

Purification of Recombinant Telomerase and Telomerase Subunit Proteins

In addition to the method of purification described in Example 3 below, it is contemplated that additional methods of purifying recombinantly produced telomerase or telomerase protein subunits will be used. For example, host cells transformed with a nucleotide sequence encoding telomerase or telomerase subunit protein(s) may be cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing the telomerase or subunit protein encoding sequence can be designed with signal sequences which direct secretion of the telomerase or telomerase subunit protein through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may join the sequence encoding the telomerase or subunit protein to a nucleotide sequence encoding a polypeptide domain.

Telomerase or telomerase subunit protein(s) may also be expressed as recombinant proteins with one or more additional polypeptide domains added to facilitate protein purification. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as polyhistidine tracts and histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle WA). The inclusion of a cleavable linker sequences such as Factor Xa or enterokinase (Invitrogen, San Diego CA) between the purification domain and telomerase or telomerase protein subunits is useful to facilitate purification. One such expression vector provides for expression of a fusion protein comprising the sequence encoding telomerase or telomerase protein subunits and nucleic acid sequence encoding 6 histidine residues followed by thioredoxin and an enterokinase cleavage site. The histidine residues facilitate purification while the enterokinase cleavage site provides a means for purifying the telomerase or telomerase protein subunit from the fusion protein. Literature pertaining to vectors containing fusion proteins is available in the art (*See e.g.,* Kroll *et al.,* DNA Cell. Biol., 12:441-53 [1993]).

In addition to recombinant production, fragments of telomerase subunit protein may be produced by direct peptide synthesis using solid-phase techniques (*See e.g.,* Merrifield, J. Am. Chem. Soc., 85:2149 [1963]). *In vitro* protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer, Foster City CA) in accordance

with the instructions provided by the manufacturer. Various fragments of telomere protein subunit may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

Uses of Telomerase and Telomerase Subunit Proteins

The rationale for use of the nucleotide and peptide sequences disclosed herein is based in part on the homology between the *E. aediculatus* telomerase 123 kDa protein subunit, the yeast protein L8543.12 (Est2), *Schizosaccharomyces*, and the human motifs observed during the development of the present invention. In particular, the yeast and 123 kDa protein contain the reverse transcriptase motif in their C-terminal regions, they share similarity in regions outside the reverse transcriptase motif, they are similarly basic (with a pI of 10.1 for the 123 kDa protein, and of 10.0 for the yeast), and they are both large (123 kDa and 103 kDa). Furthermore, in view of the reverse transcriptase motifs, these subunits are believed to comprise the catalytic core of their respective telomerases. Indeed, the reverse transcriptase motifs of the 123 kDa *E. aediculatus* telomerase protein subunit is shown in the present invention to be useful for the identification of similar sequences in other organisms.

As *E. aediculatus* and *S. cerevisiae* are so phylogenetically distant, it is contemplated that this homology provides a strong basis for predicting that human and other telomerases will contain a protein that is large, basic, and includes such reverse transcriptase motifs. Indeed, motifs have been identified within a clone encoding the human homolog of the telomerase protein. It is further contemplated that this protein is essential for human telomerase catalytic activity. This observation should prove valuable for amplification of the human telomerase gene by PCR or other methods, for screening for telomerase sequences in human and other animals, as well as for prioritizing candidate telomerase proteins or genes identified by genetic, biochemical, or nucleic acid hybridization methods. It is also contemplated that the telomerase proteins of the present invention will find use in tailing DNA 3' ends *in vitro*.

It is contemplated that expression of telomerase and/or telomerase subunit proteins in cell lines will find use in the development of diagnostics for tumors and aging factors. The nucleotide sequence may be used in hybridization or PCR technologies to diagnose the induced expression of messenger RNA sequences early in the disease process. Likewise the protein can be used to produce antibodies useful in ELISA assays or a derivative diagnostic format. Such diagnostic tests may allow different classes of human tumors or other cell-

proliferative diseases to be distinguished and thereby facilitate the selection of appropriate treatment regimens.

It is contemplated that the finding of the reverse transcriptase motifs in the telomerase proteins of the present invention will be used to develop methods to test known and yet to be described reverse transcriptase inhibitors, including nucleosides, and non-nucleosides for anti-telomerase activity.

It is contemplated that the amino acid sequence motifs disclosed herein will lead to the development of drugs (*e.g.,* telomerase inhibitors) useful in humans and/or other animals, that will arrest cell division in cancers or other disorders characterized by proliferation of cells. It is also contemplated that the telomerase proteins will find use in methods for targeting and directing RNA or RNA-tethered drugs to specific sub-cellular compartments such as the nucleus or sub-nuclear organelles, or to telomeres.

In one embodiment of the diagnostic method of the present invention, normal or standard values for telomerase mRNA expression are established as a baseline. This can be accomplished by a number of assays such as quantitating the amount of telomerase mRNA in tissues taken from normal subjects, either animal or human, with nucleic probes derived from the telomerase or telomerase protein subunit sequences provided herein (either DNA or RNA forms) using techniques which are well known in the art (*e.g.*, Southern blots, Northern blots, dot or slot blots). The standard values obtained from normal samples may be compared with values obtained from samples from subjects potentially affected by disease (*e.g.*, tumors or disorders related to aging). Deviation between standard and subject values can establish the presence of a disease state. In addition, the deviation can indicate, within a disease state, a particular clinical outcome (*e.g.*, metastatic or non-metastatic).

The nucleotide sequence encoding telomerase or telomerase protein subunits is useful when placed in an expression vector for making quantities of protein for therapeutic use. The antisense nucleotide sequence of the telomerase gene is potentially useful in vectors designed for gene therapy directed at neoplasia including metastases. Additionally, the inhibition of telomerase expression may be useful in detecting the development of disturbances in the aging process or problems occurring during chemotherapy . Alternatively, the telomerase or telomerase protein subunit encoding nucleotide sequences may used to direct the expression of telomerase or subunits in situations where it is desirable to increase the amount of telomerase activity.

Telomere Subunit Protein Antibodies

It is contemplated that antibodies directed against the telomerase subunit proteins will find use in the diagnosis and treatment of conditions and diseases associated with expression of telomerase (including the over-expression and the absence of expression). Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by a Fab expression library. Given the phylogenetic conservation of the reverse transcriptase motif in the 123 kDa subunit of the *Euplotes* telomerase, it is contemplated that antibodies directed against this subunit may be useful for the identification of homologous subunits in other organisms, including humans. It is further contemplated that antibodies directed against the motifs provided in the present invention will find use in treatment and/or diagnostic areas.

Telomerase subunit proteins used for antibody induction need not retain biological activity; however, the protein fragment, or oligopeptide must be immunogenic, and preferably antigenic. Peptides used to induce specific antibodies may have an amino acid sequence consisting of at least five amino acids, preferably at least 10 amino acids. Preferably, they should mimic a portion of the amino acid sequence of the natural protein and may contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of telomerase subunit protein amino acids may be fused with those of another protein such as keyhole limpet hemocyanin and antibody produced against the chimeric molecule. Complete telomerase used for antibody induction can be produced by co-expression of protein and RNA components in cells, or by reconstitution *in vitro* from components separately expressed or synthesized.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, etc may be immunized by injection with telomerase protein, protein subunit, or any portion, fragment or oligopeptide which retains immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants are commercially available, and include but are not limited to Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. BCG (Bacillus Calmette-Guerin) and *Corynebacterium parvum* are potentially useful adjuvants.

Monoclonal antibodies to telomerase or telomerase protein subunits be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include but are not limited to the hybridoma technique originally described by Koehler and Milstein (Koehler and Milstein, Nature 256:495-497 [1975]), the

human B-cell hybridoma technique (Kosbor *et al.*, Immunol. Today 4:72 [1983]; Cote *et al.*, Proc. Natl. Acad. Sci., 80:2026-2030 [1983]) and the EBV-hybridoma technique (Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy*, Alan R Liss Inc, New York NY, pp 77-96 [1985]).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in Orlandi *et al.* (Orlandi *et al.*,Proc. Natl. Acad. Sci., 86: 3833 [1989]; and Winter and Milstein, Nature 349:293 [1991]).

Antibody fragments which contain specific binding sites for telomerase or telomerase protein subunits may also be generated. For example, such fragments include, but are not limited to, the $F(ab')_2$ fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse *et al.*, Science 256:1275 [1989]).

A variety of protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the formation of complexes between telomerase or telomerase protein subunit and its specific antibody and the measurement of complex formation. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two noninterfering epitopes on a specific telomerase protein subunit is preferred in some situations, but a competitive binding assay may also be employed (*See e.g.*, Maddox *et al.*, J. Exp. Med., 158:1211 [1983]).

Peptides selected from the group comprising the sequences shown in Figure 32 are used to generate polyclonal and monoclonal antibodies specifically directed against human and other telomerase proteins. The peptides are useful for inhibition of protein-RNA, protein-protein interaction within the telomerase complex, and protein-DNA interaction at telomeres. Antibodies produced against these peptides are then used in various settings, including but not limited to anti-cancer therapeutics capable of inhibiting telomerase activity, for purification of native telomerase for therapeutics, for purification and cloning other components of human telomerase and other proteins associated with human telomerase, and diagnostic reagents.

Diagnostic Assays Using Telomerase Specific Antibodies

Particular telomerase and telomerase protein subunit antibodies are useful for the diagnosis of conditions or diseases characterized by expression of telomerase or telomerase protein subunits, or in assays to monitor patients being treated with telomerase, its fragments, agonists or inhibitors (including antisense transcripts capable of reducing expression of telomerase). Diagnostic assays for telomerase include methods utilizing the antibody and a label to detect telomerase in human body fluids or extracts of cells or tissues. The polypeptides and antibodies of the present invention may be used with or without modification. Frequently, the polypeptides and antibodies will be labeled by joining them, either covalently or noncovalently, with a reporter molecule. A wide variety of reporter molecules are known, several of which were described above. In particular, the present invention is useful for diagnosis of human disease, although it is contemplated that the present invention will find use in the veterinary arena.

A variety of protocols for measuring telomerase protein(s) using either polyclonal or monoclonal antibodies specific for the respective protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on the telomerase proteins or a subunit is preferred, but a competitive binding assay may be employed. These assays are described, among other places, in Maddox (Maddox *et al.*, J. Exp. Med., 158:1211 [1983]).

In order to provide a basis for diagnosis, normal or standard values for human telomerase expression must be established. This is accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with antibody to telomerase or telomerase subunit(s) under conditions suitable for complex formation which are well known in the art. The amount of standard complex formation may be quantified by comparing various artificial membranes containing known quantities of telomerase protein, with both control and disease samples from biopsied tissues. Then, standard values obtained from normal samples may be compared with values obtained from samples from subjects potentially affected by disease (*e.g.*, metastases). Deviation between standard and subject values establishes the presence of a disease state.

Drug Screening

Telomerase or telomerase subunit proteins or their catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening therapeutic compounds in any of a variety of drug screening techniques. The fragment employed in such a test may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between telomerase or the subunit protein and the agent being tested, may be measured.

Another technique for drug screening which may be used for high throughput screening of compounds having suitable binding affinity to the telomerase or telomerase protein subunit is described in detail in "Determination of Amino Acid Sequence Antigenicity" by Geysen, (Geysen, WO Application 84/03564, published on September 13, 1984, incorporated herein by reference). In summary, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with fragments of telomerase or telomerase protein subunits and washed. Bound telomerase or telomerase protein subunit is then detected by methods well known in the art. Substantially purified telomerase or telomerase protein subunit can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of binding telomerase or subunit protein(s) specifically compete with a test compound for binding telomerase or the subunit protein. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with the telomerase or subunit protein.

Uses of the Polynucleotides Encoding Telomerase Subunit Proteins

A polynucleotide sequence encoding telomerase subunit proteins or any part thereof may be used for diagnostic and/or therapeutic purposes. For diagnostic purposes, the sequence encoding telomerase subunit protein of this invention may be used to detect and quantitate gene expression of the telomerase or subunit protein. The diagnostic assay is useful to distinguish between absence, presence, and excess expression of telomerase, and to monitor regulation of telomerase levels during therapeutic intervention. Included in the scope of the invention are oligonucleotide sequences, antisense RNA and DNA molecules, and PNAs.

Another aspect of the subject invention is to provide for hybridization or PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding telomerase subunit proteins or closely related molecules. The specificity of the probe, whether it is made from a highly specific region (*e.g.*, 10 unique nucleotides in the 5' regulatory region), or a less specific region (*e.g.*, especially in the 3' region), and the stringency of the hybridization or amplification (maximal, high, intermediate or low) will determine whether the probe identifies only naturally occurring telomerase, telomerase subunit proteins or related sequences.

Probes may also be used for the detection of related sequences and should preferably contain at least 50% of the nucleotides from any of these telomerase subunit protein sequences. The hybridization probes of the subject invention may be derived from the nucleotide sequence provided by the present invention (*e.g.*, SEQ ID NO:1, 3, 62, 66, or 68), or from genomic sequence including promoter, enhancer elements and introns of the naturally occurring sequence encoding telomerase subunit proteins. Hybridization probes may be labeled by a variety of reporter groups, including commercially available radionuclides such as ^{32}P or ^{35}S, or enzymatic labels such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Other means for producing specific hybridization probes for DNAs include the cloning of nucleic acid sequences encoding telomerase subunit proteins or derivatives into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides.

Diagnostic Use

Polynucleotide sequences encoding telomerase may be used for the diagnosis of conditions or diseases with which the abnormal expression of telomerase is associated. For example, polynucleotide sequences encoding human telomerase may be used in hybridization or PCR assays of fluids or tissues from biopsies to detect telomerase expression. The form of such qualitative or quantitative methods may include Southern or northern analysis, dot blot or other membrane-based technologies; PCR technologies; dip stick, pin, chip and ELISA technologies. All of these techniques are well known in the art and are the basis of many commercially available diagnostic kits.

The human telomerase-encoding nucleotide sequences disclosed herein provide the basis for assays that detect activation or induction associated with disease (including metastasis); in addition, the lack of expression of human telomerase may be detected using the human and other telomerase-encoding nucleotide sequences disclosed herein. The nucleotide sequence may be labeled by methods known in the art and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After an incubation period, the sample is washed with a compatible fluid which optionally contains a dye (or other label requiring a developer) if the nucleotide has been labeled with an enzyme. After the compatible fluid is rinsed off, the dye is quantitated and compared with a standard. If the amount of dye in the biopsied or extracted sample is significantly elevated over that of a comparable control sample, the nucleotide sequence has hybridized with nucleotide sequences in the sample, and the presence of elevated levels of nucleotide sequences encoding human telomerase in the sample indicates the presence of the associated disease. Alternatively, the loss of expression of human telomerase sequences in a tissue which normally expresses telomerase sequences indicates the presence of an abnormal or disease state.

Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regime in animal studies, in clinical trials, or in monitoring the treatment of an individual patient. In order to provide a basis for the diagnosis of disease, a normal or standard profile for human telomerase expression must be established. This is accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with human telomerase or a portion thereof, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained for normal subjects with a dilution series of human telomerase run in the same experiment where a known amount of substantially purified human telomerase is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients affected by telomerase-associated diseases. Deviation between standard and subject values establishes the presence of disease.

Once disease is established, a therapeutic agent is administered and a treatment profile is generated. Such assays may be repeated on a regular basis to evaluate whether the values in the profile progress toward or return to the normal or standard pattern. Successive treatment profiles may be used to show the efficacy of treatment over a period of several days or several months.

PCR, which may be used as described in US Patent Nos. 4,683,195, 4,683,202, and 4,965,188 (herein incorporated by reference) provides additional uses for oligonucleotides based upon the sequence encoding telomerase subunit proteins. Such oligomers are generally chemically synthesized, but they may be generated enzymatically or produced from a recombinant source. Oligomers generally comprise two nucleotide sequences, one with sense orientation (5'→3') and one with antisense (3'← 5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantitation of closely related DNA or RNA sequences.

Additionally, methods which may be used to quantitate the expression of a particular molecule include radiolabeling (Melby *et al.*, J. Immunol. Meth., 159:235-44 [1993]) or biotinylating [Duplaa *et al.*, Anal. Biochem., 229-36 [1993]) nucleotides, co-amplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated. Quantitation of multiple samples may be speeded up by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation. A definitive diagnosis of this type may allow health professionals to begin aggressive treatment and prevent further worsening of the condition. Similarly, further assays can be used to monitor the progress of a patient during treatment. Furthermore, the nucleotide sequences disclosed herein may be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that are currently known such as the triplet genetic code, specific base pair interactions, and the like.

Therapeutic Use

Based upon its homology to other telomerase sequences, the polynucleotide encoding human telomerase disclosed herein may be useful in the treatment of metastasis; in particular, inhibition of human telomerase expression may be therapeutic. Expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences (sense or antisense) to the targeted organ, tissue or cell population. Methods which are well known to those skilled in the art can be used to construct recombinant vectors which will express antisense of the sequence encoding human telomerase. See, for example, the techniques described in Sambrook *et al.* (*supra*) and Ausubel *et al.* (*supra*).

The polynucleotides comprising full length cDNA sequence and/or its regulatory elements enable researchers to use the sequence encoding human telomerase, including the various motifs as an investigative tool in sense (Youssoufian and Lodish, Mol. Cell. Biol., 13:98-104 [1993]) or antisense (Eguchi *et al.*, Ann. Rev. Biochem., 60:631-652 [1991]) regulation of gene function. Such technology is now well known in the art, and sense or antisense oligomers, or larger fragments, can be designed from various locations along the coding or control regions.

Genes encoding human telomerase can be turned off by transfecting a cell or tissue with expression vectors which express high levels of a desired telomerase fragment. Such constructs can flood cells with untranslatable sense or antisense sequences. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until all copies are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing antisense molecules, DNA, RNA or PNA, to the control regions of the sequence encoding human telomerase (*i.e.*, the promoters, enhancers, and introns). Oligonucleotides derived from the transcription initiation site, (e.g., between -10 and +10 regions of the leader sequence) are preferred. The antisense molecules may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing compromises the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules (for a review of recent therapeutic advances using triplex DNA, see Gee *et al.*, *in* Huber and Carr, *Molecular and Immunologic Approaches*, Futura Publishing Co, Mt Kisco NY [1994]).

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Within the scope of the invention are engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of the sequence encoding human telomerase.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of

between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Antisense molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of RNA molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding human telomerase and/or telomerase protein subunits. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly can be introduced into cell lines, cells or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine and wybutosine as well as acetyl-, methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Methods for introducing vectors into cells or tissues include those methods discussed *infra*, and which are equally suitable for *in vivo, in vitro* and *ex vivo* therapy. For *ex vivo* therapy, vectors are introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient is presented in US Patent Nos. 5,399,493 and 5,437,994, the disclosure of which is herein incorporated by reference. Delivery by transfection and by liposome are quite well known in the art.

Furthermore, the nucleotide sequences encoding the various telomerase proteins and subunits disclosed herein may be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including but not limited to such properties as the triplet genetic code and specific base pair interactions.

Detection and Mapping of Related Polynucleotide Sequences in Other Genomes

The nucleic acid sequence encoding *E. aediculatus*, *S. cerevisiae*, *S. pombe*, and human telomerase subunit proteins and sequence variants thereof, may also be used to generate hybridization probes for mapping the naturally occurring homologous genomic sequence in the human and other genomes. The sequence may be mapped to a particular chromosome or to a specific region of the chromosome using well known techniques. These include *in situ* hybridization to chromosomal spreads, flow-sorted chromosomal preparations, or artificial chromosome constructions such as yeast artificial chromosomes, bacterial artificial chromosomes, bacterial P1 constructions or single chromosome cDNA libraries as reviewed by Price (Price, Blood Rev., 7:127 [1993]) and Trask (Trask, Trends Genet 7:149 [1991]).

The technique of fluorescent *in situ* hybridization (FISH) of chromosome spreads has been described, among other places, in Verma *et al.* (Verma *et al., Human Chromosomes: A Manual of Basic Techniques*, Pergamon Press, New York NY [1988]). Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of the sequence encoding human telomerase on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with the disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers are invaluable in extending genetic maps (*See e.g.,* Hudson *et al.*, Science 270:1945 [1995]). Often the placement of a gene on the chromosome of another mammalian species such as mouse (Whitehead Institute/MIT Center for Genome Research, Genetic Map of the Mouse, Database Release 10, April 28, 1995) may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques.

Pharmaceutical Compositions

The present invention also relates to pharmaceutical compositions which may comprise telomerase and/or or telomerase subunit nucleotides, proteins, antibodies, agonists, antagonists, or inhibitors, alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. Any of these molecules can be administered to a patient alone, or in combination with other agents, drugs or hormones, in pharmaceutical compositions where it is mixed with suitable excipient(s), adjuvants, and/or pharmaceutically acceptable carriers. In one embodiment of the present invention, the pharmaceutically acceptable carrier is pharmaceutically inert.

Administration Of Pharmaceutical Compositions

Administration of pharmaceutical compositions is accomplished orally or parenterally. Methods of parenteral delivery include topical, intra-arterial (*e.g.,* directly to the tumor), intramuscular, subcutaneous, intramedullary, intrathecal, intraventricular, intravenous, intraperitoneal, or intranasal administration. In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and other compounds that facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of "Remington's Pharmaceutical Sciences" (Maack Publishing Co, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, etc., suitable for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable additional compounds, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers include, but are not limited to sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; and gums including arabic and tragacanth; as well as proteins such as gelatin and collagen. If desired,

disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores are provided with suitable coatings such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound (*i.e.*, dosage).

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders such as lactose or starches, lubricants such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations for parenteral administration include aqueous solutions of active compounds. For injection, the pharmaceutical compositions of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

Manufacture And Storage

The pharmaceutical compositions of the present invention may be manufactured in a manner that known in the art (*e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes).

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic,

succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents that are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder in 1 mM-50 mM histidine, 0.1%-2% sucrose, 2%-7% mannitol at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions comprising a compound of the invention formulated in a acceptable carrier have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of human telomerase proteins, such labeling would include amount, frequency and method of administration.

Therapeutically Effective Dose

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays or in any appropriate animal model. The animal model is also used to achieve a desirable concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of protein or its antibodies, antagonists, or inhibitors which ameliorate the symptoms or condition. Therapeutic efficacy and toxicity of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals (*e.g.*, ED_{50}, the dose therapeutically effective in 50% of the population; and LD_{50}, the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD_{50}/ED_{50}. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage is chosen by the individual physician in view of the patient to be treated. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Additional factors which may be taken into account

include the severity of the disease state (*e.g.*, tumor size and location; age, weight and gender of the patient; diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy). Long acting pharmaceutical compositions might be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation. Guidance as to particular dosages and methods of delivery is provided in the literature (*See*, US Patent Nos. 4,657,760; 5,206,344; and 5,225,212, herein incorporated by reference). Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

It is contemplated, for example, that human telomerase can be used as a therapeutic molecule combat disease (*e.g.*, cancer) and/or problems associated with aging. It is further contemplated that antisense molecules capable of reducing the expression of human telomerase or telomerase protein subunits can be as therapeutic molecules to treat tumors associated with the aberrant expression of human telomerase. Still further it is contemplated that antibodies directed against human telomerase and capable of neutralizing the biological activity of human telomerase may be used as therapeutic molecules to treat tumors associated with the aberrant expression of human telomerase and/or telomerase protein subunits.

EXPERIMENTAL

The following examples are provided in order to demonstrate and further illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

In the experimental disclosure which follows, the following abbreviations apply: eq (equivalents); M (Molar); μM (micromolar); N (Normal); mol (moles); mmol (millimoles); μmol (micromoles); nmol (nanomoles); g (grams); mg (milligrams); μg (micrograms); ng (nanograms); l or L (liters); ml (milliliters); μl (microliters); cm (centimeters); mm (millimeters); μm (micrometers); nm (nanometers); °C (degrees Centigrade); RPN (ribonucleoprotein); remN (2'-*O*-methylribonucleotides); dNTP (deoxyribonucleotide); dH_2O (distilled water); DDT (dithiothreitol); PMSF (phenylmethylsulfonyl fluoride); TE (10 mM Tris HCl, 1 mM EDTA, approximately pH 7.2); KGlu (potassium glutamate); SSC (salt and sodium citrate buffer); SDS (sodium dodecyl sulfate); PAGE (polyacrylamide gel

electrophoresis); Novex (Novex, San Diego, CA); BioRad (Bio-Rad Laboratories, Hercules, CA); Pharmacia (Pharmacia Biotech, Piscataway, NJ); Boehringer-Mannheim (Boehringer-Mannheim Corp., Concord, CA); Amersham (Amersham, Inc., Chicago, IL); Stratagene (Stratagene Cloning Systems, La Jolla, CA); NEB (New England Biolabs, Beverly, MA); Pierce (Pierce Chemical Co., Rockford, IL); Beckman (Beckman Instruments, Fullerton, CA); Lab Industries (Lab Industries, Inc., Berkeley, CA); Eppendorf (Eppendorf Scientific, Madison, WI); and Molecular Dynamics (Molecular Dynamics, Sunnyvale, CA).

EXAMPLE 1

Growth of *Euplotes aediculatus*

In this Example, cultures of *E. aediculatus* were obtained from Dr. David Prescott, MCDB, University of Colorado. Dr. Prescott originally isolated this culture from pond water, although this organism is also available from the ATCC (ATCC #30859). Cultures were grown as described by Swanton *et al.,* (Swanton *et al.,*, Chromosoma 77:203 [1980]), under non-sterile conditions, in 15-liter glass containers containing *Chlorogonium* as a food source. Organisms were harvested from the cultures when the density reached approximately 10^4 cells/ml.

EXAMPLE 2

Preparation of Nuclear Extracts

In this Example, nuclear extracts of *E. aediculatus* were prepared using the method of Lingner *et al.,* (Lingner *et al.,* Genes Develop., 8:1984 [1994]), with minor modifications, as indicated below. Briefly, cells grown as described in Example 1 were concentrated with 15 μm Nytex filters and cooled on ice. The cell pellet was resuspended in a final volume of 110 ml TMS/PMSF/spermidinephosphate buffer. The stock TMS/PMSF/spermidine phosphate buffer was prepared by adding 0.075 g spermidine phosphate (USB) and 0.75 ml PMSF (from 100 mM stock prepared in ethanol) to 150 ml TMS. TMS comprised 10 mM Tris-acetate, 10 mM $MgCl_2$, 85.5752 g sucrose/liter, and 0.33297 g $CaCl_2$/liter, pH 7.5.

After resuspension in TMS/PMSF/spermidinephosphate buffer, 8.8 ml 10% NP-40 and 94.1 g sucrose were added and the mixture placed in a siliconized glass beaker with a

stainless steel stirring rod attached to an overhead motor. The mixture was stirred until the cells were completely lysed (approximately 20 minutes). The mixture was then centrifuged for 10 minutes at 7500 rpm (8950 x g), at 4°C, using a Beckman JS-13 swing-out rotor. The supernatant was removed and nuclei pellet was resuspended in TMS/PMSF/spermidine phosphate buffer, and centrifuged again, for 5 minutes at 7500 rpm (8950 x g), at 4°C, using a Beckman JS-13 swing-out rotor.

The supernatant was removed and the nuclei pellet was resuspended in a buffer comprised of 50 mM Tris-acetate, 10 mM $MgCl_2$, 10% glycerol, 0.1% NP-40, 0.4 M KGlu, 0.5 mM PMSF, pH 7.5, at a volume of 0.5 ml buffer per 10 g of harvested cells. The resuspended nuclei were then dounced in a glass homogenizer with approximately 50 strokes, and then centrifuged for 25 minutes at 14,000 rpm at 4°C, in an Eppendorf centrifuge. The supernatant containing the nuclear extract was collected, frozen in liquid nitrogen, and stored at -80°C until used.

EXAMPLE 3

Purification of Telomerase

In this Example, nuclear extracts prepared as described in Example 2 were used to purify *E. aediculatus* telomerase. In this purification protocol, telomerase was first enriched by chromatography on an Affi-Gel-heparin column, and then extensively purified by affinity purification with an antisense oligonucleotide. As the template region of telomerase RNA is accessible to hybridization in the telomerase RNP particle, an antisense oligonucleotide (*i.e.*, the "affinity oligonucleotide") was synthesized that was complementary to this template region as an affinity bait for the telomerase. A biotin residue was included at the 5' end of the oligonucleotide to immobilize it to an avidin column.

Following the binding of the telomerase to the oligonucleotide, and extensive washing, the telomerase was eluted by use of a displacement oligonucleotide. The affinity oligonucleotide included DNA bases that were not complementary to the telomerase RNA 5' to the telomerase-specific sequence. As the displacement oligonucleotide was complementary to the affinity oligonucleotide for its entire length, it was able to form a more thermodynamically stable duplex than the telomerase bound to the affinity oligonucleotide.

Thus, addition of the displacement oligonucleotide resulted in the elution of the telomerase from the column.

In this Example, the nuclear extracts prepared from 45 liter cultures were frozen until a total of 34 ml of nuclear extract was collected. This corresponded to 630 liters of culture (*i.e.*, approximately 4 x 10^9 cells). The nuclear extract was diluted with a buffer to 410 ml, to provide final concentrations of 20 mM Tris-acetate, 1 mM $MgCl_2$, 0.1 mM EDTA, 33 mM KGlu, 10% (vol/vol) glycerol, 1 mM dithiothreitol (DTT), and 0.5 mM phenylmethylsulfonyl fluoride (PMSF), at a pH of 7.5.

The diluted nuclear extract was applied to an Affi-Gel-heparin gel column (Bio-Rad), with a 230 ml bed volume and 5 cm diameter, equilibrated in the same buffer and eluted with a 2-liter gradient from 33 to 450 mM KGlu. The column was run at 4°C, at a flow rate of 1 column volume/hour. Fractions of 50 mls each were collected and assayed for telomerase activity as described in Example 4. Telomerase was eluted from the column at approximately 170 mM KGlu. Fractions containing telomerase (approximately 440 ml) were pooled and adjusted to 20 mM Tris-acetate, 10 mM $MgCl_2$, 1 mM EDTA, 300 mM KGlu, 10% glycerol, 1 mM DTT, and 1% Nonidet P-40. This buffer was designated as "WB."

To this preparation, 1.5 nmol of each of two competitor DNA oligonucleotides (5'-TAGACCTGTTAGTGTACATTTGAATTGAAGC-3' (SEQ ID NO:28)) and (5'-TAGACCTGTTAGGTTGGATTTGTGGCATCA-3' (SEQ ID NO:29)), 50 μg yeast RNA (Sigma), and 0.3 nmol of biotin-labelled telomerase-specific oligonucleotide (5'-biotin-TAGACCTGTTA-(rmeG)$_2$-(rmeU)$_4$-(rmeG)$_4$-(rmeU)$_4$-remG-3')(SEQ ID NO:60), were added per ml of the pool. The 2-*O*-methyribonucleotides of the telomerase specific oligonucleotides were complementary to the telomerase RNA template region; the deoxyribonucleotides were not complementary. The inclusion of competitor, non-specific DNA oligonucleotides increased the efficiency of the purification, as the effects of nucleic acid binding proteins and other components in the mixture that would either bind to the affinity oligonucleotide or remove the telomerase from the mixture were minimized.

This material was then added to Ultralink immobilized neutravidin plus (Pierce) column material, at a volume of 60 μl of suspension per ml of pool. The column material was pre-blocked twice for 15 minutes each blocking, with a preparation of WB containing 0.01% Nonidet P-40, 0.5 mg BSA, 0.5 mg/ml lysozyme, 0.05 mg/ml glycogen, and 0.1 mg/ml yeast RNA. The blocking was conducted at 4°C, using a rotating wheel to thoroughly

block the column material. After the first blocking step, and before the second blocking step, the column material was centrifuged at 200 x g for 2 minutes to pellet the matrix.

The pool-column mixture was incubated for 8 minutes at 30°C, and then for an additional 2 hours at 4°C, on a rotating wheel (approximately 10 rpm; Labindustries) to allow binding. The pool-column mixture was then centrifuged 200 xg for 2 minutes, and the supernatant containing unbound material was removed. The pool-column mixture was then washed. This washing process included the steps of rinsing the pool-column mixture with WB at 4°C, washing the mixture for 15 minutes with WB at 4°C, rinsing with WB, washing for 5 minutes at 30°C, with WB containing 0.6 M KGlu, and no Nonidet P-40, washing 5 minutes at 25°C with WB, and finally, rinsing again with WB. The volume remaining after the final wash was kept small, in order to yield a ratio of buffer to column material of approximately 1:1.

Telomerase was eluted from the column material by adding 1 nmol of displacement deoxyoligonucleotide (5'-$CA_4C_4A_4C_2TA_2CAG_2TCTA$-3')(SEQ ID NO:30), per ml of column material and incubating at 25°C for 30 minutes. The material was centrifuged for 2 minutes 14,000 rpm in a microcentrifuge (Eppendorf), and the eluate collected. The elution procedure was repeated twice more, using fresh displacement oligonucleotide each time. As mentioned above, because the displacement oligonucleotide was complementary to the affinity oligonucleotide, it formed a more thermodynamically stable complex with the affinity oligonucleotide than the telomerase. Thus, addition of the displacement oligonucleotide to an affinity-bound telomerase resulted in efficient elution of telomerase under native conditions. The telomerase appeared to be approximately 50% pure at this stage, as judged by analysis on a protein gel. The affinity purification of telomerase and elution with a displacement oligonucleotide is shown in Figure 1 (panels A and B, respectively). In this Figure, the 2'-*O*-methyl sugars of the affinity oligonucleotide are indicated by the bold line. The black and shaded oval shapes in this Figure are intended to graphically represent the protein subunits of the present invention.

The protein concentrations of the extract and material obtained following Affi-Gel-heparin column chromatography, were determined using the method of Bradford (Bradford, Anal. Biochem., 72:248 [1976]), using BSA as the standards. Only a fraction of the telomerase preparation was further purified on a glycerol gradient.

The sedimentation coefficient of telomerase was determined by glycerol gradient centrifugation, as described in Example 8.

Table 1 below is a purification table for telomerase purified according to the methods of this Example. The telomerase was enriched 12-fold in nuclear extracts, as compared to whole cell extracts, with a recovery of 80%; 85% of telomerase was solubilized from nuclei upon extraction.

Table 1. Purification of Telomerase

Fraction	Protein (mg)	Telomerase (pmol of RNP)	Telomerase/Protein/pmol of RNP/mg	Recovery (%)	Purification Factor
Nuclear Extract	2020	1720	0.9	100	1
Heparin	125	1040	8.3	60	10
Affinity	0.3**	680	2270	40	2670
Glycerol Gradient	NA*	NA*	NA*	25	NA*

*NA=Not available

**This value was calculated from the measured amount of telomerase (680 pmol), by assuming a purity of 50% (based on a protein gel).

EXAMPLE 4

Telomerase Activity

At each step in the purification of telomerase, the preparation was analyzed by three separate assays, one of which was activity, as described in this Example. In general, telomerase assays were done in 40 µl containing 0.003-0.3 µl of nuclear extract, 50 mM Tris-Cl (pH 7.5), 50 mM KGlu, 10 mM $MgCl_2$, 1 mM DTT, 125 µM dTTP, 125 µM dGTP, and approximately 0.2 pmoles of 5'-^{32}P-labelled oligonucleotide substrate (*i.e.*, approximately 400,000 cpm). Oligonucleotide primers were heat-denatured prior to their addition to the reaction mixture. Reactions were assembled on ice and incubated for 30 minutes at 25°C. The reactions were stopped by addition of 200 µl of 10 mM Tris-Cl (pH 7.5), 15 mM EDTA, 0.6% SDS, and 0.05 mg/ml proteinase K, and incubated for at least 30 minutes at 45°C. After ethanol precipitation, the products were analyzed on denaturing 8% PAGE gels, as known in the art (*See e.g.*, Sambrook *et al.*, 1989).

EXAMPLE 5

Quantification of Telomerase Activity

In this Example, quantification of telomerase activity through the purification procedure is described. Quantitation was accomplished by assaying the elongation of oligonucleotide primers in the presence of dGTP and [α -^{32}P]dTTP. Briefly, 1 μM 5'-$(G_4T_4)_2$-3' oligonucleotide was extended in a 20 μl reaction mixture in the presence of 2 μl of [α -^{32}P]dTTP (10 mCi/ml, 400 Ci/mmol; 1 Ci=37 GBq), and 125 μM dGTP as described by (Lingner *et al.,* Genes Develop., 8:1984 [1994]), and loaded onto an 8% PAGE sequencing gel as known in the art (*See e.g.,* Sambrook *et al.,* 1989).

The results of this study are shown in Figure 3. In lane 1, there is no telomerase present (*i.e.,* a negative control); lanes 2, 5, 8, and 11 contained 0.14 fmol telomerase; lanes 3, 6, 9, and 12 contained 0.42 fmol telomerase; and lanes 4, 7, 10, and 13 contained 1.3 fmol telomerase. Activity was quantified using a PhosphorImager (Molecular Dynamics) using the manufacturer's instructions. It was determined that under these conditions, 1 fmol of affinity-purified telomerase incorporated 21 fmol of dTTP in 30 minutes.

As shown in this figure, the specific activity of the telomerase did not change significantly through the purification procedure. Affinity-purified telomerase was fully active. However, it was determined that at high concentrations, an inhibitory activity was detected and the activity of crude extracts was not linear. Thus, in the assay shown in Figure 3, the crude extract was diluted 700-7000-fold. Upon purification, this inhibitory activity was removed and no inhibitory effect was detected in the purified telomerase preparations, even at high enzyme concentrations.

EXAMPLE 6

Gel Electrophoresis and Northern Blots

As indicated in Example 4, at each step in the purification of telomerase, the preparation was analyzed by three separate assays. This Example describes the gel electrophoresis and blotting procedures used to quantify telomerase RNA present in fractions and analyze the integrity of the telomerase ribonucleoprotein particle.

Denaturing Gels and Northern Blots

In this Example, synthetic T7-transcribed telomerase RNA of known concentration served as the standard. Throughout this investigation, the RNA component was used as a measure of telomerase.

A construct for phage T7 RNA polymerase transcription of *E. aediculatus* telomerase RNA was produced, using the polymerase chain reaction (PCR). The telomerase RNA gene was amplified with primers that annealed to either end of the gene. The primer that annealed at the 5' end also encoded a hammerhead ribozyme sequence to generate the natural 5' end upon cleavage of the transcribed RNA, a T7-promoter sequence, and an *Eco*RI site for subcloning. The sequence of this 5' primer was 5'-GCGGGAATTCTAATACGACTCACTATAGGGAAGAAACTCTGATGAGGCCGAAAG GCCGAAACTCCACGAAAGTGGAGTAAGTTTCTCGATAATTGATCTGTAG-3' (SEQ ID NO:31). The 3' primer included an *Ear*I site for termination of transcription at the natural 3' end, and a *Bam*HI site for cloning. The sequence of this 3' primer was 5'-CGGGGATCCTCTTCAAAAGATGAGAGGACAGCAAAC-3' (SEQ ID NO:32). The PCR amplification product was cleaved with *Eco*RI and *Bam*HI, and subcloned into the respective sites of pUC19 (NEB), to give "pEaT7." The correctness of this insert was confirmed by DNA sequencing. T7 transcription was performed as described by Zaug *et al.,* Biochemistry 33:14935 [1994]), with *Ear*I-linearized plasmid. RNA was gel-purified and the concentration was determined (an A_{260} of 1 = 40 μg/ml). This RNA was used as a standard to determine the telomerase RNA present in various preparations of telomerase.

The signal of hybridization was proportional to the amount of telomerase RNA, and the derived RNA concentrations were consistent with, but slightly higher than those obtained by native gel electrophoresis. Comparison of the amount of whole telomerase RNA in whole cell RNA to serial dilutions of known T7 RNA transcript concentrations indicated that each *E. aediculatus* cell contained approximately 300,000 telomerase molecules.

Visualization of the telomerase was accomplished by Northern blot hybridization to its RNA component, using the methods described by Lingner *et al.* (Linger *et al.,* Genes Develop., 8:1984 [1994]). Briefly, RNA (less than or equal to 0.5 μg/lane) was resolved on an 8% PAGE and electroblotted onto a Hybond-N membrane (Amersham), as known in the art (*See e.g.,* Sambrook *et al.,* 1989). The blot was hybridized overnight in 10 ml of 4x SSC, 10x Denhardt's solution, 0.1% SDS, and 50 μg/ml denatured herring sperm DNA,. After pre-hybridizing for 3 hours, 2 x 10^6 cpm probe/ml hybridization solution was added. The

randomly labelled probe was a PCR-product that covered the entire telomerase RNA gene. The blot was washed with several buffer changes for 30 minutes in 2x SSC, 0.1% SDS, and then washed for 1 hour in 0.1x SSC and 0.1% SDS at 45°C.

Native Gels and Northern Blots

In this experiment, the purified telomerase preparation was run on native (*i.e.*, non-denaturing) gels of 3.5% polyacrylamide and 0.33% agarose, as known in the art and described by Lamond and Sproat (Lamond and Sproat, [1994], *supra*). The telomerase comigrated approximately with the xylene cyanol dye.

The native gel results indicated that telomerase was maintained as an RNP throughout the purification protocol. Figure 2 is a photograph of a Northern blot showing the mobility of the telomerase in different fractions on a non-denaturing gel as well as *in vitro* transcribed telomerase. In this figure, lane 1 contained 1.5 fmol telomerase RNA, lane 2 contained 4.6 fmol telomerase RNA, lane 3 contained 14 fmol telomerase RNA, lane 4 contained 41 fmol telomerase RNA, lane 5 contained nuclear extract (42 fmol telomerase), lane 6 contained Affi-Gel-heparin-purified telomerase (47 fmol telomerase), lane 7 contained affinity-purified telomerase (68 fmol), and lane 8 contained glycerol gradient-purified telomerase (35 fmol).

As shown in Figure 2, in nuclear extracts, the telomerase was assembled into an RNP particle that migrated slower than unassembled telomerase RNA. Less than 1% free RNA was detected by this method. However, a slower migrating telomerase RNP complex was also sometimes detected in extracts. Upon purification on the Affi-Gel-heparin column, the telomerase RNP particle did not change in mobility (Figure 2, lane 6). However, upon affinity purification the mobility of the RNA particle slightly increased (Figure 2, lane 7), perhaps indicating that a protein subunit or fragment had been lost. On glycerol gradients, the affinity-purified telomerase did not change in size, but approximately 2% free telomerase RNA was detectable (Figure 2, lane 8), suggesting that a small amount of disassembly of the RNP particle had occurred.

EXAMPLE 7

Telomerase Protein Composition

In this Example, the analysis of the purified telomerase protein composition are described.

In this Example, glycerol gradient fractions obtained from Example 8, were separated on a 4-20% polyacrylamide gel (Novex). Following electrophoresis, the gel was stained with Coomassie brilliant blue. Figure 4 shows a photograph of the gel. Lanes 1 and 2 contained molecular mass markers (Pharmacia) as indicated on the left side of the gel shown in Figure 4. Lanes 3-5 contained glycerol gradient fraction pools as indicated on the top of the gel (*i.e.*, lane 3 contained fractions 9-14, lane 4 contained fractions 15-22, and lane 5 contained fractions 23-32). Lane 4 contained the pool with 1 pmol of telomerase RNA. In lanes 6-9 BSA standards were run at concentrations indicated at the top of the gel in Figure 4 (*i.e.*, lane 6 contained 0.5 pmol BSA, lane 7 contained 1.5 pmol BSA, lane 8 contained 4.5 BSA, and lane 9 contained 15 pmol BSA).

As shown in Figure 4, polypeptides with molecular masses of 120 and 43 kDa co-purified with the telomerase. The 43 kDa polypeptide was observed as a doublet. It was noted that the polypeptide of approximately 43 kDa in lane 3 migrated differently than the doublet in lane 4; it may be an unrelated protein. The 120 kDa and 43 kDa doublet each stained with Coomassie brilliant blue at approximately the level of 1 pmol, when compared with BSA standards. Because this fraction contained 1 pmol of telomerase RNA, all of which was assembled into an RNP particle (*See*, Figure 2, lane 8), there appear to be two polypeptide subunits that are stoichiometric with the telomerase RNA. However, it is also possible that the two proteins around 43 kDa are separate enzyme subunit.s

Affinity-purified telomerase that was not subjected to fractionation on a glycerol gradient contained additional polypeptides with apparent molecular masses of 35 and 37 kDa, respectively. This latter fraction was estimated to be at least 50% pure. However, the 35 kDa and 37 kDa polypeptides that were present in the affinity-purified material were not reproducibly separated by glycerol gradient centrifugation. These polypeptides may be contaminants, as they were not visible in all activity-containing preparations.

EXAMPLE 8

Sedimentation Coefficient

The sedimentation coefficient for telomerase was determined by glycerol gradient centrifugation. In this Example, nuclear extract and affinity-purified telomerase were fractionated on 15-40% glycerol gradients containing 20 mM Tris-acetate, with 1 mM $MgCl_2$, 0.1 mM EDTA, 300 mM KGlu, and 1 mM DTT, at pH 7.5. Glycerol gradients were poured in 5 ml (13 x 51 mm) tubes, and centrifuged using an SW55Ti rotor (Beckman) at 55,000 rpm for 14 hours at 4°C.

Marker proteins were run in a parallel gradient and had a sedimentation coefficient of 7.6 S for alcohol dehydrogenase (ADH), 113 S for catalase, 17.3 S for apoferritin, and 19.3 S for thyroglobulin. The telomerase peak was identified by native gel electrophoresis of gradient fractions followed by blot hybridization to its RNA component.

Figure 5 is a graph showing the sedimentation coefficient for telomerase. As shown in this Figure, affinity-purified telomerase co-sedimented with catalase at 11.5 S, while telomerase in nuclear extracts sedimented slightly faster, peaking around 12.5 S. Therefore, consistent with the mobility of the enzyme in native gels, purified telomerase appears to have lost a proteolytic fragment or a loosely associated subunit.

The calculated molecular mass for telomerase, if it is assumed to consist of one 120 kDa protein subunit, one 43 kDa subunit, and one RNA subunit of 66 kDa, adds up to a total of 229 kDa. This is in close agreement with the 232 kDa molecular mass of catalase. However, the sedimentation coefficient is a function of the molecular mass, as well as the partial specific volume and the frictional coefficient of the molecule, both of which are unknown for the telomerase RNP.

EXAMPLE 9

Substrate Utilization

In this Example, the substrate requirements of telomerase were investigated. One simple model for DNA end replication predicts that after semi-conservative DNA replication, telomerase extends double-stranded, blunt-ended DNA molecules. In a variation of this model, a single-stranded 3' end is created by a helicase or nuclease after replication. This 3' end is then used by telomerase for binding and extension.

To determine whether telomerase is capable of elongating blunt-ended molecules, model hairpins were synthesized with telomeric repeats positioned at their 3' ends. These primer substrates were gel-purified, 5'-end labelled with polynucleotide kinase, heated at 0.4 μM to 80°C for 5 minutes, and then slowly cooled to room temperature in a heating block, to allow renaturation and helix formation of the hairpins. Substrate mobility on a non-denaturing gel indicated that very efficient hairpin formation was present, as compared to dimerization.

In this Example, assays were performed with unlabelled 125 μM dGTP, 125 μM dTTP, and 0.02 μM 5'-end-labelled primer (5'-^{32}P-labelled oligonucleotide substrate) in 10 μl reaction mixtures that contained 20 mM Tris-acetate, with 10 mM $MgCl_2$, 50 mM KGlu, and 1 mM DTT, at pH 7.5. These mixtures were incubated at 25°C for 30 minutes. Reactions were stopped by adding formamide loading buffer (*i.e.,* TBE, formamide, bromthymol blue, and cyanol, Sambrook, 1989, *supra*).

Primers were incubated without telomerase ("-"), with 5.9 fmol of affinity-purified telomerase ("+"), or with 17.6 fmol of affinity-purified telomerase ("+++"). Affinity-purified telomerase used in this assay was dialyzed with a membrane having a molecular cut-off of 100 kDa, in order to remove the displacement oligonucleotide. Reaction products were separated on an 8% PAGE/urea gel containing 36% formamide, to denature the hairpins. The sequences of the primers used in this study, as well as their lane assignments are shown in Table 2.

TABLE 2. Primer Sequences

Lane	Primer Sequence (5' to 3')	SEQ ID NO:
1-3	$C_4(A_4C_4)_3CACA(G_4T_4)_3G_4$	SEQ ID NO:33
4-6	$C_2(A_4C_4)_3CACA(G_4T_4)_3G_4$	SEQ ID NO:34
7-9	$(A_4C_4)_3CACA(G_4T_4)_3G_4$	SEQ ID NO:35
10-12	$A_2C_4(A_4C_4)_2CACA(G_4T_4)_3G_4$	SEQ ID NO:36
13-15	$C_4(A_4C_4)_2CACA(G_4T_4)_3$	SEQ ID NO:37
16-18	$(A_4C_4)_3CACA(G_4T_4)_3$	SEQ ID NO:38
19-21	$A_2C_4(A_4C_4)_2CACA(G_4T_4)_3$	SEQ ID NO:39
22-24	$C_4(A_4C_4)_2CACA(G_4T_4)_3$	SEQ ID NO:40
25-27	$C_2(A_4C_4)_2CACA(G_4T_4)_3$	SEQ ID NO:41
28-30	$(A_4C_4)_2CACA(G_4T_4)_3$	SEQ ID NO:42

The gel results are shown in Figure 6. Lanes 1-15 contained substrates with telomeric repeats ending with four G residues. Lanes 16-30 contained substrates with telomeric repeats ending with four T residues. The putative alignment on the telomerase RNA template is indicated in Figure 7 (SEQ ID NOS:43 and 44, and 45 and 46, respectively). It was assumed that the primer sets anneal at two very different positions in the template shown in Figure 7 (*i.e.,* 7A and 7B, respectively). This may have affected their binding and/or elongation rate.

Figure 8 shows a lighter exposure of lanes 25-30 in Figure 6. The lighter exposure of Figure 8 was taken in order to permit visualization of the nucleotides that are added and the positions of pausing in elongated products. Percent of substrate elongated for the third lane in each set was quantified on a PhosphorImager, as indicated on the bottom of Figure 6.

The substrate efficiencies for these hairpins were compared with double-stranded telomere-like substrates with overhangs of differing lengths. A model substrate that ended with four G residues (see lanes 1-15 of Figure 6), was not elongated when it was blunt ended (see lanes 1-3). However, slight extension was observed with an overhang length of two bases; elongation became efficient when the overhang was at least 4 bases in length. The telomerase acted in a similar manner with a double-stranded substrate that ended with four T residues, with a 6-base overhang required for highly efficient elongation. In Figure 6, the

faint bands below the primers in lanes 10-15 that are independent of telomerase represent shorter oligonucleotides in the primer preparations.

The lighter exposure of lanes 25-30 in Figure 8 shows a ladder of elongated products, with the darkest bands correlating with the putative 5' boundary of the template (as described by Lingner *et al.,* Genes Develop., 8:1984 [1994]). The abundance of products that correspond to other positions in the template suggested that pausing and/or dissociation occurs at sites other than the site of translocation with the purified telomerase.

As shown in Figure 6, double-stranded, blunt-ended oligonucleotides were not substrates for telomerase. To determine whether these molecules would bind to telomerase, a competition experiment was performed. In this experiment, 2 nM of 5'-end labelled substrate with the sequence $(G_4T_4)_2$ (SEQ ID NO:61), or a hairpin substrate with a six base overhang respectively were extended with 0.125 nM telomerase (Figure 6, lanes 25-27). Although the same unlabeled oligonucleotide substrates competed efficiently with labelled substrate for extension, no reduction of activity was observed when the double-stranded blunt-ended hairpin oligonucleotides were used as competitors, even in the presence of 100-fold excess hairpins.

These results indicated that double-stranded, blunt-ended oligonucleotides cannot bind to telomerase at the concentrations tested in this Example. Rather, a single-stranded 3' end is required for binding. It is likely that this 3' end is required to base pair with the telomerase RNA template.

EXAMPLE 10

Cloning & Sequencing of the 123 kDa Polypeptide

In this Example, the cloning of the 123 kDa polypeptide of telomerase (*i.e.*, the 123 kDa protein subunit) is described. In this study, an internal fragment of the telomerase gene was amplified by PCR, with oligonucleotide primers designed to match peptide sequences that were obtained from the purified polypeptide obtained in Example 3, above. The polypeptide sequence was determined using the nanoES tandem mass spectroscopy methods known in the art and described by Calvio *et al.*, RNA 1:724-733 [1995]). The oligonucleotide primers used in this Example had the following sequences, with positions that were degenerate shown in parentheses--5'-TCT(G/A)AA(G/A)TA(G/A)TG(T/G/A)GT(G/A/T/C)A(T/G/A)(G/A)TT(G/A)TTCAT-3' (SEQ ID NO:47), AND 5'-GCGGATCCATGAA(T/C)CC(A/T)GA(G/A)AA(T/C)CC(A/T)AA(T/C)GT-3' (SEQ ID NO:48).

A 50 µl reaction contained 0.2 mM dNTPs, 0.15 µg *E. aediculatus* chromosomal DNA, 0.5 µl *Taq* (Boehringer-Mannheim), 0.8 µg of each primer, and 1x reaction buffer (Boehringer-Mannheim). The reaction was incubated in a thermocycler (Perkin-Elmer), using the following--5 minutes at 95°C, followed by 30 cycles of 1 minute at 94°C, 1 minute at 52°C, and 2 minutes at 72°C. The reaction was completed by 10 minute incubation at 72°C.

A genomic DNA library was prepared from the chromosomal *E. aediculatus* DNA by cloning blunt-ended DNA into the *Sma*I site of pCR-Script plasmid vector (Stratagene). This library was screened by colony hybridization, with the radiolabelled, gel-purified PCR product. Plasmid DNA of positive clones was prepared and sequenced by the dideoxy method (Sanger *et al.*, Proc. Natl. Acad. Sci., 74:5463 [1977]) or manually, through use of an automated sequencer (ABI). The DNA sequence of the gene encoding this polypeptide is shown in Figure 9 (SEQ ID NO:1). The start codon in this sequence inferred from the DNA sequence, is located at nucleotide position 101, and the open reading frame ends at position 3193. The genetic code of *Euplotes* differs from other organisms in that the "UGA" codon encodes a cysteine residue. The amino acid sequence of the polypeptide inferred from the DNA sequence is shown in Figure 10 (SEQ ID NO:2), and assumes that no unusual amino acids are inserted during translation and no post-translational modification occurs.

EXAMPLE 11

Cloning & Sequencing of the 43 kDa Polypeptide

In this Example, the cloning of the 43 kDa polypeptide of telomerase (*i.e.,* the 43 kDa protein subunit) is described. In this study, an internal fragment of the telomerase gene was amplified by PCR, with oligonucleotide primers designed to match peptide sequences that were obtained from the purified polypeptide obtained in Example 3, above. The polypeptide sequence was determined using the nanoES tandem mass spectroscopy methods known in the art and described by Calvio *et al.,* RNA 1:724-733 [1995]). The oligonucleotide primers used in this Example had the following sequences--5'-NNNGTNAC(C/T/A)GG(C/T/A)AT(C/T/A)AA(C/T)AA-3' (SEQ ID NO:49), and 5'-(T/G/A)GC(T/G/A)GT(C/T)TC(T/C)TG(G/A)TC(G/A)TT(G/A)TA-3' (SEQ ID NO:50). In this sequence, "N" indicates the presence of any of the four nucleotides (*i.e.,* A, T, G, or C).

A 50 µl reaction contained 0.2 mM dNTPs, 0.2 µg *E. aediculatus* chromosomal DNA, 0.5 µl *Taq* (Boehringer-Mannheim), 0.8 µg of each primer, and 1x reaction buffer (Boehringer-Mannheim). The reaction was incubated in a thermocycler (Perkin-Elmer), using the following--5 minutes at 95°C, followed by 30 cycles of 1 minute at 94°C, 1 minute at 52°C, and 1 minutes at 72°C. The reaction was completed by 10 minute incubation at 72°C.

A genomic DNA library was prepared from the chromosomal *E. aediculatus* DNA by cloning blunt-ended DNA into the *Sma*I site of pCR-Script plasmid vector (Stratagene). This library was screened by colony hybridization, with the radiolabelled, gel-purified PCR product. Plasmid DNA of positive clones was prepared and sequenced by the dideoxy method (Sanger *et al.,* Proc. Natl. Acad. Sci., 74:5463 [1977]) or manually, through use of an automated sequencer (ABI). The DNA sequence of the gene encoding this polypeptide is shown in Figure 11 (SEQ ID NO:3). Three potential reading frames are shown for this sequence, as shown in Figure 12. For clarity, the amino acid sequence is indicated below the nucleotide sequence in all three reading frames. These reading frames are designated as "a," "b," and "c" (SEQ ID NOS:4-6). A possible start codon is encoded at nucleotide position 84 in reading frame "c." They coding region could end at position 1501 in reading frame "b."

Early stop codons, indicated by asterisks in this figure, occur in all three reading frames between nucleotide position 337-350.

The "La-domain" is indicated in bold-face type. Further downstream, the protein sequence appears to be encoded by different reading frames, as none of the three frames is uninterrupted by stop codons. Furthermore, peptide sequences from purified protein are encoded in all three frames. Therefore, this gene appears to contain intervening sequences, or in the alternative, the RNA is edited. Other possibilities include ribosomal frame-shifting or sequence errors. However, the homology to the La-protein sequence remains of significant interest. Again, in *Euplotes*, the "UGA" codon encodes a cysteine residue.

EXAMPLE 12

Amino Acid and Nucleic Acid Comparisons

In this Example, comparisons between various reported sequences and the sequences of the 123 kDa and 43 kDa telomerase subunit polypeptides were made.

Comparisons with the 123 kDa *E. aediculatus* Telomerase Subunit

The amino acid sequence of the 123 kDa *Euplotes aediculatus* polypeptide was compared with the sequence of the 80 kDa telomerase protein subunit of *Tetrahymena thermophila* (GenBank accession #U25641) in order to investigate their similarity. The nucleotide sequence as obtained from GenBank (SEQ ID NO:51) encoding this protein is shown in Figure 19. The amino acid sequence of this protein as obtained from GenBank (SEQ ID NO:52) is shown in Figure 20. The sequence comparison between the 123 kDa *E. aediculatus* and 80 kDa *T. thermophila* is shown in Figure 13. In this figure, the *E. aediculatus* sequence is the upper sequence (SEQ ID NO:2), while the *T. thermophila* sequence is the lower sequence (SEQ ID NO:52). In this Figure, as well as Figures 14-16, identities are indicated by vertical bars, while single dots between the sequences indicate somewhat similar amino acids, and double dots between the sequences indicate more similar amino acids. The observed identity was determined to be approximately 19%, while the percent similarity was approximately 45%, values similar to what would be observed with any random protein sequence.

The amino acid sequence of the 123 kDa *Euplotes aediculatus* polypeptide was also compared with the sequence of the 95 kDa telomerase protein subunit of *Tetrahymena thermophila* (GenBank accession #U25642), in order to investigate their similarity. The nucleotide sequence as obtained from GenBank (SEQ ID NO:53) encoding this protein is shown in Figure 21. The amino acid sequence of this protein as obtained from GenBank (SEQ ID NO:54) is shown in Figure 22. This sequence comparison is shown in Figure 14. In this figure, the *E. aediculatus* sequence is the upper sequence (SEQ ID NO:2), while the *T. thermophila* sequence is the lower sequence (SEQ ID NO:54); identities are indicated by vertical bars. The observed identity was determined to be approximately 20%, while the percent similarity was approximately 43%, values similar to what would be observed with any random protein sequence.

Significantly, the amino acid sequence of the 123 kDa *E. aediculatus* polypeptide contains the five motifs (SEQ ID NOS:13 and 18) characteristic of reverse transcriptases. The 123 kDa polypeptide was also compared with the polymerase domains of various reverse transcriptases (SEQ ID NOS:14-17, and 19-22). Figure 17 shows the alignment of the 123 kDa polypeptide with the putative yeast homolog (L8543.12 or ESTp)(SEQ ID NOS: 17 and 22). The amino acid sequence of L8543.12 (or ESTp) obtained from GenBank is shown in Figure 23 (SEQ ID NO:55).

Four motifs (A, B, C, and D) were included in this comparison. In this Figure 17, highly conserved residues are indicated by white letters on a black background. Residues of the *E. aediculatus* sequences that are conserved in the other sequence are indicated in bold; the "h" indicates the presence of a hydrophobic amino acid. The numerals located between amino acid residues of the motifs indicates the length of gaps in the sequences. For example, the "100" shown between motifs A and B reflects a 100 amino acid gap in the sequence between the motifs.

Genbank searches identified a yeast protein (Genbank accession #u20618), and gene "L8543.12" (Est2), containing amino acid sequence that shows some homology to the *E. aediculatus* 123 kDa telomerase subunit. Based on the observations that both proteins contain reverse transcriptase motifs in their C-terminal regions; both proteins share similarity in regions outside the reverse transcriptase motif; the proteins are similarly basic (pI = 10.1 for *E. aediculatus* and pI=10.0 for the yeast); and both proteins are large (123 kDa for *E. aediculatus* and 103 kDa for the yeast), these sequences comprise the catalytic core of their respective telomerases. It is contemplated that based on this observation of homology in two phylogenetically distinct organisms as *E. aediculatus* and yeast, the human telomerase will

contain a protein that has the same characteristics (*i.e.*, reverse transcriptase motifs, is basic, and large [> 100 kDa]).

Comparisons with the 43 kDa *E. aediculatus* Telomerase Subunit

The amino acid sequence of the "La-domain" of the 43 kDa *Euplotes aediculatus* polypeptide was compared with the sequence of the 95 kDa telomerase protein subunit of *Tetrahymena thermophila* (described above) in order to investigate their similarity. This sequence comparison is shown in Figure 15. In this figure, the *E. aediculatus* sequence is the upper sequence (SEQ ID NO:9), while the *T. thermophila* sequence is the lower sequence (SEQ ID NO:10); identities are indicated by vertical bars. The observed identity was determined to be approximately 23%, while the percent similarity was approximately 46%, values similar to what would be observed with any random protein sequence.

The amino acid sequence of the "La-domain" of the 43 kDa *Euplotes aediculatus* polypeptide was compared with the sequence of the 80 kDa telomerase protein subunit of *Tetrahymena thermophila* (described above) in order to investigate their similarity. This sequence comparison is shown in Figure 16. In this figure, the *E. aediculatus* sequence is the upper sequence (SEQ ID NO:11), while the *T. thermophila* sequence is the lower sequence (SEQ ID NO:12); identities are indicated by vertical bars. The observed identity was determined to be approximately 26%, while the percent similarity was approximately 49%, values similar to what would be observed with any random protein sequence.

The amino acid sequence of a domain of the 43 kDa *E. aediculatus* polypeptide (SEQ ID NO:23) was also compared with La proteins from various other organisms (SEQ ID NOS:24-27). These comparisons are shown in Figure 18. In this Figure, highly conserved residues are indicated by white letters on a black background. Residues of the *E. aediculatus* sequences that are conserved in the other sequence are indicated in bold.

EXAMPLE 13

Identification of Telomerase Protein Subunits in Another Organism

In this Example, the sequences identified in the previous Examples above, were used to identify the telomerase protein subunits of *Oxytricha trifallax*, a ciliate that is very distantly related to *E. aediculatus*. In this Example, primers were chosen based on the conserved region of the *E. aediculatus* 123 kDa polypeptide which comprised the reverse transcriptase domain motifs. Suitable primers were synthesized and used in a PCR reaction with total DNA from *Oxytricha*. The *Oxytricha* DNA was prepared according to methods known in the art. The PCR products were then cloned and sequenced using methods known in the art.

The oligonucleotide sequences used as the primers were as follows: 5'-(T/C)A(A/G)AC(T/A/C)AA(G/A)GG(T/A/C)AT(T/C)CC(C/T/A)(C/T)A(G/A)GG-3' (SEQ ID NO:56) and 5'-(G/A/T)GT(G/A/T)ATNA(G/A)NA(G/A)(G/A)TA(G/A)TC(G/A)TC-3' (SEQ ID NO:57). Positions that were degenerate are shown in parenthesis, with the alternative bases shown within the parenthesis. "N" represents any of the four nucleotides.

In the PCR reaction, a 50 µl reaction contained 0.2 mM dNTPs, 0.3 µg *Oxytricha trifallax* chromosomal DNA, 1 µl *Taq* polymerase (Boehringer-Mannheim), 2 micromolar of each primer, 1x reaction buffer (Boehringer-Mannheim). The reaction was incubated in a thermocycler (Perkin-Elmer) under the following conditions: 1x 5 min at 95°C, 30 cycles consisting of 1 min at 94°C, 1 min at 53°C, and 1 min at 72°C, followed by 1x 10 min at 72°C. The PCR-product was gel-purified and sequenced by the dideoxy-method, by methods known well in the art (*e.g.,* Sanger *et al.,* Proc. Natl. Acad. Sci. 74, 5463-5467 (1977).

The deduced amino acid sequence of the PCR product was determined and compared with the *E. aediculatus* sequence. Figure 24 shows the alignment of these sequences, with the *O. trifallax* sequence (SEQ ID NO:58) shown in the top row, and the *E. aediculatus* sequence (SEQ ID NO:59) shown in the bottom row. As can be seen from this Figure, there is a great deal of homology between the *O. trifallax* polypeptide sequence identified in this Example with the *E. aediculatus* polypeptide sequence. Thus, it is clear that the sequences identified in the present invention are useful for the identification of homologous telomerase

protein subunits in other eukaryotic organisms. Indeed, development of the present invention has identified homologous telomerase sequences in multiple, diverse species.

EXAMPLE 15

Identification of *Tetrahymena* Telomerase Sequences

In this Example, a *Tetrahymena* clone was produced that shares homology with the *Euplotes* sequences, and EST2p.

This experiment utilized PCR with degenerate oligonucleotide primers directed against conserved motifs to identify regions of homology between *Tetrahymena*, *Euplotes*, and EST2p sequences. The PCR method used in this Example is a novel method that is designed to specifically amplify rare DNA sequences from complex mixtures. This method avoids the problem of amplification of DNA products with the same PCR primer at both ends (*i.e.*, single primer products) commonly encountered in PCR cloning methods. These single primer products produce unwanted background and can often obscure the amplification and detection of the desired two-primer product. The method used in these experiment preferentially selects for two-primer products. In particular, one primer is biotinylated and the other is not. After several rounds of PCR amplification, the products are purified using streptavidin magnetic beads and two primer products are specifically eluted using heat denaturation. This method finds use in settings other than the experiments described in this Example. Indeed, this method finds use in application in which it is desired to specifically amplify rare DNA sequences, including the preliminary steps in cloning methods such as 5' and 3; RACE, and any method that uses degenerate primers in PCR.

A first PCR run was conducted using *Tetrahymena* template macronuclear DNA isolated using methods known in the art, and the 24-mer forward primer with the sequence 5' biotin-GCCTATTT(TC)TT(TC)TA(TC)(GATC)(GATC)(GATC)AC(GATC)GA-3' (SEQ ID NO:70) designated as "K231," corresponding to the FFYXTE region (SEQ ID NO:71), and the 23-mer reverse primer with the sequence 5'-CCAGATAT(GATC)A(TGA)(GATC)A(AG)(AG)AA(AG)TC(AG)TC- 3' (SEQ ID NO:72), designated as "K220," corresponding to the DDFL(FIL)I region (SEQ ID NO:73). This PCR reaction contained 2.5 µl DNA (50 ng), 4 µl of each primer (20 µM), 3 µl 10x PCR buffer,

3 µl 10x dNTPs, 2 µl Mg, 0.3 µl *Taq*, and 11.2 µl dH_2O. The mixture was cycled for 8 cycles of 94°C for 45 seconds, 37°C for 45 seconds, and 72 °C for 1 minute.

This PCR reaction was bound to 200 µl streptavidin magnetic beads, washed with 200 µl TE, resuspended in 20 µl dH_2O and then heat-denatured by boiling at 100°C for 2 minutes. The beads were pulled down and the eluate removed. Then, 2.5 µl of this eluate was subsequently reamplified using the above conditions, with the exception bieng that 0.3 µl of α -^{32}P dATP was included, and the PCR was carried out for 33 cycles. This reaction was run a 5% denaturing polyacrylamide gel, and the appropriate region was cut out of the gel. These products were then reamplified for an additional 34 cycles, under the conditions listed above, with the exception being that a 42°C annealing temperature was used.

A second PCR run was conducted using *Tetrahymena* macronuclear DNA template isolated using methods known in the art, and the 23-mer forward primer with the sequence 5' ACAATG(CA)G(GATC)(TCA)T(GATC)(TCA)T(GATC)CC(GATC)AA(AG)AA-3' (SEQ ID NO:74), designated as "K228," corresponding to the region R(LI)(LI)PKK (SEQ ID NO:75), and a reverse primer with the sequence 5'-ACGAATC(GT)(GATC)GG(TAG)AT(GATC)(GC)(TA)(AG)TC(AG)TA(AG)CA 3' (SEQ ID NO:76), designated "K224," corresponding to the CYDSIPR region (SEQ ID NO:77). This PCR reaction contained 2.5 µl DNA (50 ng), 4 µl of each primer (20 µM), 3 µl 10x PCR buffer, 3 µl 10x dNTPs, 2 µl Mg, 0.3 µl α -^{32}P dATP, 0.3 µl *Taq*, and 10.9 µl dH_2O. This reaction was run on a 5% denaturing polyacrylamide gel, and the appropriate region was cut out of the gel. These products were reamplified for an additional 34 cycles, under the conditions listed above, with the exception being that a 42°C annealing temperature was used.

Ten µl of the reaction product from run 1 were bound to streptavidin-coated magnetic beads in 200 µl TE. The beads were washed with 200 µl TE, and then then resuspended in 20 µl of dH_2O, heat denatured, and the eluate was removed. Next, 2.5 µl of this eluate was reamplified for 33 cycles using the conditions indicated above. The reaction product from run 2 was then added to the beads and diluted with 30 µl 0.5x SSC. The mixture was heated from 94°C to 50°C. The eluate was removed and the beads were washed three times in 0.5x SSC at 55°C. The beads were then resuspended in 20 µl dH_2O, heat denatured, and the eluate was removed, designated as "round 1 eluate" and saved.

To isolate the *Tetrahymena* band, the round 1 eluate was reamplified with the forward primer K228 (SEQ ID NO:74) and reverse primer K227 (SEQ ID NO:78) with the sequence

5'- CAATTCTC(AG)TA(AG)CA(GATC)(CG)(TA)(CT)TT(AGT)AT(GA)TC-3' (SEQ ID NO:78), corresponding to the DIKSCYD region (SEQ ID NO:79). The PCR reactions were conducted as described above. The reaction products were run on a 5% polyacrylamide gel; the band corresponding to approximately 295 nucleotides was cut from the gel and sequenced.

The clone designated as 168-3 was sequenced. The DNA sequence (including the primer sequences) was found to be:

GATTACTCCCGAAGAAAGGATCTTTCCGTCCAATCATGACTTTCTTAAGAAAGGA
CAAGCAAAAAAATATTAAGTTAAATCTAAATTAAATTCTAATGGATAGCCAACTT
GTGTTTAGGAATTTAAAAGACATGCTGGGATAAAAGATAGGATACTCAGTCTTTG
ATAATAAACAAATTTCAGAAAAATTTGCCTAATTCATAGAGAAATGGAAAAATA
AAGGAAGACCTCAGCTATATTATGTCACTCTAGACATAAAGACTTGCTAC (SEQ ID NO:80).

Additional sequence of this gene was obtained by PCR using one unique primer designed to match the sequence from 168-3 ("K297" with the sequence 5'-GAGTGACATAATATACGTGA-3'; SEQ ID NO:111), and the K231 (FFYXTE) primer. The sequence of the fragment obtained from this reaction, together with 168-3 is as follows (without the primer sequences):

AAACACAAGGAAGGAAGTCAAATATTCTATTACCGTAAACCAATATGGAAATTA
GTGAGTAAATTAACTATTGTCAAAGTAAGAATTTAGTTTTCTGAAAAGAATAAAT
AAATGAAAAATAATTTTTATCAAAAAATTTAGCTTGAAGAGGAGAATTTGGAAA
AAGTTGAAGAAAAATTGATACCAGAAGATTCATTTTAGAAATACCCTCAAGGAA
AGCTAAGGATTATACCTAAAAAAGGATCTTTCCGTCCAATCATGACTTTCTTAAG
AAAGGACAAGCAAAAAAATATTAAGTTAAATCTAAATTAAATTCTAATGGATAG
CCAACTTGTGTTTAGGAATTTAAAAGACATGCTGGGATAAAAGATAGGATACTC
AGTCTTTGATAATAAACAAATTTCAGAAAAATTTGCCTAATTCATAGAGAAATGG
AAAAATAAAGGAAGACCTCAGCTATATTATGTCACTCTA (SEQ ID NO:81).

The amino acid sequence corresponding to this DNA fragment was found to be:

KHKEGSQIFYYRKPIWKLVSKLTIVKVRIQFSEKNKQMKNNFYQKIQLEEENLEKVEE
KLIPEDSFQKYPQGKLRIIPKKGSFRPIMTFLRKDKQKNIKLNLNQILMDSQLVFRNLK
DMLGQKIGYSVFDNKQISEKFAQFIEKWKNKGRPQLYYVTL (SEQ ID NO:82).

This amino acid sequence was then aligned with other telomerase genes (EST2p, and *Euplotes*). The alignment is shown in Figure 31. Consensus sequence is also shown in this Figure.

EXAMPLE 16

Identification of *Schizosaccharomyces pombe* Telomerase Sequences

In this Example, the *tez1* sequence of *S. pombe* was identified as a homolog of the *E. aediculatus* p123, and *S. cerevisiae* Est2p.

Figure 33 provides an overall summary of these experiments. In this Figure, the top portion (Panel A) shows the relationship of two overlapping genomic clones, and the 5825 bp portion that was sequenced. The region designated at "*tez1*$^{+}$" is the protein coding region, with the flanking sequences indicated as well, the box underneath the 5825 bp region is an approximately 2 kb *Hin*dIII fragment that was used to make tez1 disruption construct, as described below.

The bottom half of Figure 33 (Panel B) is a "close-up" schematic of this same region of DNA. The sequence designated as "original PCR" is the original degenerate PCR fragment that was generated with degenerate oligonucleotide primer pair designed based on *Euplotes* sequence motif 4 (B') and motif 5 (C), as described in previous Examples.

PCR With Degenerate Primers

PCR using degenerate primers was used to find the homolog of the *E. aediculatus* p123 in *S. pombe.* Figure 34 shows the sequences of the degenerate primers (designated as "poly 4" and "poly 1") used in this reaction. The PCR runs were conducted using the same buffer as described in previous Examples (*See e.g.,* Example 10, above), with a 5 minute ramp time at 94°C, followed by 30 cycles of 94°C for 30 seconds, 50°C for 45 seconds, and 72°C for 30 seconds, and 7 minutes at 72°C, followed by storage at 4°C. PCR runs were conducted using varied conditions, (*i.e.,* various concentrations of *S. pombe* DNA and $MgCl_2$ concentrations). The PCR products were run on agarose gels and stained with ethidium bromide as described above. Several PCR runs resulted in the production of three bands (designated as "T," "M," and "B"). These bands were re-amplified and run on gels using the same conditions as described above. Four bands were observed following this re-amplification ("T," "M1," "M2," and "B"), as shown in Figure 35. These four bands were then re-amplified using the same conditions as described above. The third band from the top of the lane in Figure 35 was identified as containing the correct sequence for telomerase

protein. The PCR product designated as M2 was found to show a reasonable match with other telomerase proteins, as indicated in Figure 36. In addition to the alignment shown, this Figure also shows the actual sequence of *tez1*. In this Figure, the asterisks indicate residues shared with all four sequences (*Oxytricha* "Ot"; *E. aediculatus* "Ea_p123"; *S. cerevisiae* "Sc_p103"; and M2), while the circles (*i.e.,* dots) indicate similar amino acid residues.

3' RT PCR

In order to obtain additional sequence information, 3' and 5' RT PCR were conducted on the telomerase candidate identified in Figure 36. Figure 37 provides a schematic of the 3' RT PCR strategy used. First, cDNA was prepared from mRNA using the oligonucleotide primer "Q_T," (5'-CCA GTG AGC AGA GTG ACG AGG ACT CGA GCT CAA GCT TTT TTT TTT TTT TT-3'; SEQ ID NO:102), then using this cDNA as a template for PCR with "Q_O" (5'-CCA GTG AGC AGA GTG ACG-3'; SEQ ID NO:103), and a primer designed based on the original degenerated PCR reaction (*i.e.,* "M2-T" with the sequence 5'-G TGT CAT TTC TAT ATG GAA GAT TTG ATT GAT G-3' (SEQ ID NO:109). The second PCR reaction (*i.e.,* nested PCR)
with "Q_I" (5'-GAG GAC TCG AGC TCA AGC-3'; SEQ ID NO:104), and another PCR primer designed with sequence derived from the original degenerate PCR reaction or "M2-T2" with the sequence 5'-AC CTA TCG TTT ACG AAA AAG AAA GGA TCA GTG-3'; SEQ ID NO:110). The buffers used in this PCR were the same as described above, with amplification conducted beginning with a ramp up of 94° for 5 min, followed by 30 cycles of 94° for 30 sec, 55°C for 30 sec, and 72°C for 3 min), followed by 7 minutes at 72°C. The reaction products were stored at 4°C until use.

Screening of Genomic and cDNA Libraries

After obtaining this extra sequence information, several genomic and cDNA libraries were screened to identify any libraries that contain this telomerase candidate gene. The approach used, as well as the libraries and results are shown in Figure 38. In this Figure, Panel A lists the libraries tested in this experiment; Panel B shows the regions used; Panels C and D show the dot blot hybridization results obtained with these libraries. Positive libraries were then screened by colony hybridization to obtain genomic and cDNA version of *tez1* gene. In this experiment, approximately 3×10^4 colonies from the *Hin*dIII genomic library were screened and six positive clones were identified (approximately 0.01%). DNA was then

prepared from two independent clones (A5 and B2). Figure 39 shows the results obtained with the *Hin*dIII-digested A5 and B2 positive genomic clones.

In addition, cDNA REP libraries were used. Approximately 3 x 10^5 colonies were screened, and 5 positive clones were identified (0.002%). DNA was prepared from three independent clones (2-3, 4-1, and 5-20). In later experiments, it was determined that 2-3 and 5-20 contained identical inserts.

5' RT PCR

As the cDNA version of gene produced to this point was not complete, 5' RT-PCR was conducted in order to obtain a full length clone. The strategy is schematically shown in Figure 40. In this experiment, cDNA was prepared using DNA oligonucleotide primer "M2-B" (5'-CAC TGA TCC TTT CTT TTT CGT AAA CGA TAG GT-3'; SEQ ID NO:105) and "M2-B2" (5'-C ATC AAT CAA ATC TTC CAT ATA GAA ATG ACA-3'; SEQ ID NO:106), designed from known regions of *tez1* identified previously. An oligonucleotide linker PCR Adapt SfiI with a phosphorylated 5' end ("P") (P-GGG CCG TGT TGG CCT AGT TCT CTG CTC-3'; SEQ ID NO:107) was then ligated at the 3' end of this cDNA, and this construct was used as the template for nested PCR. In the first round of PCR, PCR Adapt SFI and M2-B were used as the primers; while PCR Adapt SfiII (5-GAG GAG GAG AAG AGC AGA GAA CTA GGC CAA CAC GCC CC-3'; SEQ ID NO:108), and M2-B2 (5'- ATC AAT CAA ATC TTC CAT ATA GAA ATG ACA-3'; SEQ ID NO:106) were used as primers in the second round. Nested PCR was used to increase specificity of reaction.

Sequence Alignments

Once the sequence of *tez1* was identified, it was compared with sequences previously described. Figure 41 shows the alignment of reverse transcriptase (RT) domains from telomerase catalytic subunits of *S. pombe* ("S.p. Tez1p"), *S. cerevisiae* ("S.c. Est2p"), and *E. aediculatus* p123 ("E.a. p123"). In this Figure, "h" indicates hydrophobic residues, while "p" indicates small polar residues, and "c" indicates charged residues. The amino acid residues indicated above the alignment shows the consensus RT motif of Y. Xiong and T.H. Eickbush (Y. Xiong and T.H. Eickbush, EMBO J., 9: 3353-3362 [1990]). The asterisks indicate the residues that are conserved for all three proteins. "Motif O" is identified herein as a motif specific to this telomerase subunit and not found in reverse transcriptases in general. It is therefore valuable in identifying other amino acid sequences as being good candidates for telomerase catalytic subunits.

Figure 42 shows the alignment of entire sequences from *Euplotes* ("Ea_p123"), S. *cerevisiae* ("Sc_Est2p"), and *S. pombe* ("Sp_Tez1p"). In Panel A, the shaded areas indicate residues shared between two sequences. In Panel B, the shaded areas indicate residues shared between all three sequences.

Genetic Disruption of tez1

In this Example, the effects of disruption of *tez1* were investigated. As telomerase is involved in telomere maintenance, it was hypothesized that if *tez1* were indeed a telomerase component, disruption of *tez1* was expected to cause gradual telomere shortening.

In these experiments, homologous recombination was used to specifically disrupt the *tez1* gene in *S. pombe*. This approach is schematically illustrated in Figure 43. As indicated in Figure 43, wild type *tez1* was replaced with a fragment containing the *ura4* or *LEU2* marker.

The disruption of tez1 gene was confirmed by PCR (Figure 44) , and Southern blot was performed to check for telomere length. Figure 45 shows the Southern blot results for this experiment. Because an *Apa I* restriction enzyme site is present immediately adjacent to telomeric sequence in *S. pombe*, digestion of *S. pombe* genomic DNA preparations permits analysis of telomere length. Thus, DNA from *S. pombe* was digested with *ApaI* and the digestion products were run on an agarose gel and probed with a telomeric sequence-specific probe to determine whether the telomeres of disrupted *S. pombe* cells were shortened. The results are shown in Figure 45. From these results, it was clear that disruption of the *tez1* gene caused a shortening of the telomeres.

EXAMPLE 17

Cloning and Characterization of Human Telomerase Protein and cDNA

In this Example, the nucleic and amino acid sequence information for human telomerase was determined. Partial homologous sequences were first identified in a BLAST search conducted using the *Euplotes* 123 kDa peptide and nucleic acid sequences, as wells as *Schizosaccharomyces* protein and corresponding cDNA (tez1) sequences. The human sequences (also referred to as "hTCP1.1") were identified from a partial cDNA clone

(GenBank accession #AA281296). Sequences from this clone were aligned with the sequences determined as described in previous Examples.

Figure 25 shows the sequence alignment of the *Euplotes* ("p123"), *Schizosaccharomyces* ("tez1"), Est2p (*i.e.,* the *S. cerevisiae* protein encoded by the *Est2* nucleic acid sequence, and also referred to herein as "L8543.12"), and the human homolog identified in this comparison search. The amino acid sequence of this aligned portion is provided in SEQ ID NO:67 (the cDNA sequence is provided in SEQ ID NO:62), while the portion of tez1 shown in Figure 25 is provided in SEQ ID NO:63. The portion of Est2 shown in this Figure is also provided in SEQ ID NO:64, while the portion of p123 shown is also provided in SEQ ID NO:65. Figure 29 shows the amino acid sequence of tez1 (SEQ ID NO:69), while Figure 30 shows the DNA sequence of *tez1* (SEQ ID NO:68). In Figure 30, the introns and other non-coding regions, are shown in lower case, while the exons (*i.e.,* coding regions) are shown in upper case.

As shown in Figure 25, there are regions that are highly conserved among these proteins. For example, as shown in this Figure, there are regions of identity in "Motif 0," "Motif 1, "Motif 2," and "Motif 3." The identical amino acids are indicated with an asterisk (*), while the similar amino acid residues are indicated by a circle (●). This indicates that there are regions within the telomerase motifs that are conserved among a wide variety of eukaryotes, ranging from yeast to ciliates, to humans. It is contemplated that additional organisms will likewise contain such conserved regions of sequence. Figure 27 shows the partial amino acid sequence of the clone encoding human telomerase motifs (SEQ ID NO:67), while Figure 28 shows the corresponding DNA sequence of the Genbank #AA281296 clone.

Sanger dideoxy sequencing and other methods were used, as known in the art to obtain complete sequence information of the Genbank clone #AA281296. Some of the primers used in the sequencing are shown in Table 3. These primers were designed to hybridize to the clone (GenBank accession #AA281296), based on sequence complementarity to either plasmid backbone sequence or the sequence of the human cDNA insert in the clone.

Table 3. Primers

Primer	Sequence	SEQ ID NO:
TCP1.1	GTGAAGGCACTGTTCAGCG	SEQ ID NO:87
TCP1.2	GTGGATGATTTCTTGTTGG	SEQ ID NO:88
TCP1.3	ATGCTCCTGCGTTTGGTGG	SEQ ID NO:89
TCP1.4	CTGGACACTCAGCCCTTGG	SEQ ID NO:90
TCP1.5	GGCAGGTGTGCTGGACACT	SEQ ID NO:91
TCP1.6	TTTGATGATGCTGGCGATG	SEQ ID NO:92
TCP1.7	GGGGCTCGTCTTCTACAGG	SEQ ID NO:93
TCP1.8	CAGCAGGAGGATCTTGTAG	SEQ ID NO:94
TCP1.9	TGACCCCAGGAGTGGCACG	SEQ ID NO:95
TCP1.10	TCAAGCTGACTCGACACCG	SEQ ID NO:96
TCP1.11	CGGCGTGACAGGGCTGC	SEQ ID NO:97
TCP1.12	GCTGAAGGCTGAGTGTCC	SEQ ID NO:98
TCP1.13	TAGTCCATGTTCACAATCG	SEQ ID NO:99

From these experiments, it was determined that the EcoRI-NotI insert of the Genbank #AA281296 clone contains only a partial open reading frame for the human telomerase protein, although it may encode an active fragment of that protein. The open reading frame in the clone encodes an approximately 63 kD protein. The sequence of the longest open reading frame identified is shown in Figure 47 (SEQ ID NO:100). The ORF begins at the ATG codon with the "met" indicated in the Figure. The poly A tail at the 3' end of the sequence is also shown. Figure 48 shows a tentative alignment of telomerase reverse transcriptase proteins from the human sequence (human Telomerase Core Protein 1, "Hs TCP1"), *E. aediculatus* p123 ("Ep p123), *S. pombe* tez1 ("Sp Tez1"), *S. cerevisiae* EST2 (Sc Est2"), and consensus sequence. In this Figure various motifs are indicated.

To obtain a full-length clone, probing of a cDNA library and 5 '-RACE were used to obtain clones encoding portions of the previously uncloned regions. In these experiments, RACE (Rapid Amplification of cDNA Ends; *See e.g., M.A.* Frohman, "RACE: Rapid Amplification of cDNA Ends," *in* Innis *et al.* (eds), *PCR Protocols: A Guide to Methods and Applications* [1990], pp. 28-38; and Frohman *et al.*, Proc. Natl. Acad. Sci., 85:8998-9002 [1988]) was used to generate material for sequence analysis. Four such clones were generated and used to provide additional 5' sequence information (pFWRP5, 6, 19, and 20).

In addition, human cDNA libraries (inserted into lambda) were probed with the EcoRI-NotI fragment of the clone (#AA281296). One lambda clone, designated "lambda 25-1.1," (ATCC accession #_____) was identified as containing complementary sequences. Figure 54 shows a restriction map of this lambda clone. The human cDNA insert from this clone was subcloned as an *Eco*RI restriction fragment into the *Eco*RI site of commercially available phagemid pBluescriptIISK+ (Stratagene), to create the plasmid "pGRN121," which was deposited with the ATCC (ATCC accession #209016). Preliminary results indicated that plasmid pGRN121 contains the entire open reading frame (ORF) sequence encoding the human telomerase protein.

The cDNA insert of plasmid pGRN121 was sequenced using techniques known in the art. Figure 49 provides a restriction site and function map of plasmid pGRN121 identified based on this preliminary work. The results of this preliminary sequence analysis are shown in Figure 50. From this analysis, and as shown in Figure 49, a putative start site for the coding region was identified at approximately 50 nucleotides from the *Eco*RI site (located at position 707), and the location of the telomerase-specific motifs, "FFYVTE" (SEQ ID NO:112), "PKP," "AYD," "QG", and "DD," were identified, in addition to a putative stop site at nucleotide #3571 (*See,* Figure 51). Figure 51 shows the DNA and corresponding amino acid sequences for the open reading frames in the sequence ("a" [SEQ ID NOS:174-201], "b" [SEQ ID NOS:202-214], and "c" [SEQ ID NOS:215-223]). However, due to the preliminary nature of the early sequencing work, the reading frames for the various motifs were found not to be in alignment.

Additional analysis conducted on the pGRN121 indicated that the plasmid contained significant portions from the 5'-end of the coding sequence not present on the Genbank accession #AA281296 clone. Furthermore, pGRN121 was found to contain a variant coding sequence that includes an insert of approximately 182 nucleotides. This insert was found to be absent from the Genbank accession #AA281296 clone. As with the *E. aediculatus*

sequences, such variants can be tested in functional assays, such as telomerase assays to detect the presence of functional telomerase in a sample.

Further sequence analysis resolved the cDNA sequence of pGRN121, to provide a contiguous open reading frame that encodes a protein of molecular weight of approximately 127,000 daltons, and 1132 amino acids as shown in Figure 53 (SEQ ID NOS:224 and 225). A refined map of pGRN121 based on this analysis, is provided in Figure 52.

From the above, it is clear that the present invention provides nucleic acid and amino acid sequences, as well as other information regarding telomerase, telomerase protein subunits, and motifs from various organisms, in addition to methods for identification of homologous structures in other organisms in addition to those described herein.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.